PROGRESS IN BIOPHYSICS

AND BIOPHYSICAL CHEMISTRY

6

OL.
6
956

Editor
J. A. V. BUTLER
Professor of Physical Chemistry
University of London
Institute of Cancer Research
Royal Cancer Hospital, London

PERGAMON PRESS
OXFORD · LONDON · NEW YORK · PARIS

PERGAMON PRESS LTD.
Headington Hill Hall, Oxford
4 & 5 Fitzroy Square, London W.1

PERGAMON PRESS INC.
122 East 55th Street, New York 22, N.Y.
1404 New York Avenue N.W., Washington 5 D.C.

PERGAMON PRESS S.A.R.L.
24 Rue des Écoles, Paris Ve

PERGAMON PRESS G.m.b.H.
Kaiserstrasse 75, Frankfurt am Main

First Published 1956
Second Impression 1961

Originally set and printed in Great Britain
Reprinted by photolithography in Hungary, 1961.

VOI
6
19

CONTENTS

		PAGE
	PREFACE	*vii*
	ACKNOWLEDGEMENTS	*viii*
1	PROTAMINES AND NUCLEOPROTAMINES K. FELIX, H. FISCHER and A. KREKELS, Institut fur Vegetative Physiologie, Weigerstrasse, Frankfurt-a-M, Sud., 10, Germany	1
2	THE STRUCTURE OF CHROMOSOMES E. J. AMBROSE, Chester Beatty Research Institute, The Royal Cancer Hospital, Fulham Road, London, S.W.3	25
3	THE DONNAN EQUILIBRIUM J. TH. OVERBEEK, Van't Hoff Laboratorium, Sterrenbos 19, Utrecht, Holland	57
4	BIOLOGY AND BIOPHYSICAL PROPERTIES OF TRANSFORMING PRINCIPLES STEPHEN ZAMENHOF, College of Physicians and Surgeons, Columbia University, New York 21, N.Y., U.S.A.	85
5	BIOPHYSICAL ASPECTS OF NEURO-MUSCULAR TRANSMISSION J. DEL CASTILLO and B. KATZ, University College, Gower Street, London, W.C.1	121
6	MODELS FOR BIOLOGICAL EXCITATION PROCESSES U. F. FRANCK, Max-Planck-Institut für Physikalische Chemie, Bunsenstrasse 10, Göttingen, Germany	171
7	PHYSICAL ASPECTS OF THE SENSE ORGANS HL. DE VRIES, Natuurkundig Laboratorium der Rijks-Universiteit, Westersingel 34, Groningen, Holland	207
	NAME INDEX	265
	SUBJECT INDEX	270

VOL.
6
956

VOL
6
195

PREFACE

THE appearance of the sixth volume of this series may be a convenient point to look back to see how far the aims of this series, as stated in the original Preface, have been achieved. As was pointed out there, since each volume could only deal at length with a few topics, it would necessarily be some time before many of the aspects of this field of science could receive attention. Although some 46 articles have now been printed, the attentive reader will have no difficulty in discovering important subjects which have not yet been dealt with. We hope that future volumes will do something to fill some of these gaps. Volumes of this kind depend on the willingness of authors who are authorities on their subject to devote much time and trouble to review an extensive body of recent research. The editors have been fortunate in obtaining the co-operation of many outstanding scientists, and they hope that this will be continued.

An attempt has been made to keep a reasonable balance between the different aspects of the subject and to include the work of those who are primarily physicists or physical chemists and those who are biologists. It will have been apparent that there is still something of a gap between the two approaches, although it is rather rapidly diminishing and it is hoped that these volumes will assist in promoting its early disappearance.

I regret to announce that, with the completion of Volume V, my colleague, Professor J. T. RANDALL, found it necessary to retire from the Editorship owing to the increasing pressure of other duties. The present volume, for which I am responsible, is a transitorial one and I am happy to state that, beginning with Volume VII, which is in active preparation, Professor B. KATZ, F.R.S., of the Department of Biophysics, University College, London, has agreed to join me as an Editor.

J. A. V. BUTLER

ACKNOWLEDGEMENTS

THE authors and publishers wish to thank the many scientists and authorities concerned for permission to use previously published illustrations. The relevant source is indicated by a reference number or quoted in the caption to the illustration.

VOL
6
195

NAME INDEX

Abderhalden, E. 10
Adair, L. S. 80
Adair, M. E. 80
Akabori, Sh. 11
Alexander, H. E. 89, 90, 91, 92, 93, 94, 95, 106, 108, 109
Alfert, M. 43
Allmand, A. J. 61, 63
Alloway, J. L. 88
Ambrose, E. J. 29, 30, 31, 32, 33, 36, 37, 38, 39, 40, 45, 46
Ando, T. 6, 10
Astbury, W. T. 29, 31
Atlas, S. M. 101
Austrian, R. 90, 91
Autrum, H. 218, 249, 253
Avery, O. T. 88, 90, 93, 97, 103, 104, 105, 107, 108, 109, 110

Babcock, K. L. 62, 75
Bahr, C. F. 38
Balbianni, E. G. 34
Bamford, C. H. 29
Barlow, H. B. 246
Barnes, T. C. 152
Beermann, W. 39, 41, 42
Békésy, G. von 226, 227, 228, 229, 234, 236
Bergmann, M. 11
Bernstein, M. H. 47
Berry, G. P. 88
Bethe, A. 184, 201
Beutner, R. 152
Birbeck, M. S. C. 38
Bird, M. J. 45
Blair, E. A. 203
Bleeker, J. D. J. W. 231
Block, R. J. 4, 5
Bohr, G. F. 42
Boivin, A. 43, 89
Bolling, D. 4
Boltzmann 65, 67
Bonhoeffer, K. F. 173, 189, 191, 203
Bouman, M. A. 256, 258
Brachet, J. 36
Branson, H. R. 29, 30
Bray, C. W. 229
Brockmann, H. 10
Brown, G. L. 139
Brown, L. 29
Brücke, E. 251
Bruins, H. R. 73
Buck, J. B. 39
Burke, W. 166
Butler, J. A. V. 26, 28, 30, 31, 32, 33, 37, 47, 52

Callan, H. G. 42, 47
Caspersson, T. 35, 36
Castillo, J. del 127, 134, 135, 139, 140, 142, 145, 147, 148, 153, 161, 162, 163, 167
Chapman, D. L. 68
Chargaff, E. 16, 28
Charlwood, P. A. 80
Claude, A. 38
Collie, B. 81
Conway, B. E. 26, 28, 30
Corey, R. B. 29, 30
Corfield, M. C. 5
Crampton, C. F. 27, 28
Crawford, B. H. 251
Crick, F. H. C. 16, 28, 30, 31, 44, 45, 49
Croft, H. J. 27
Cuckow, F. W. 38, 39, 45

Dackerman, M. G. 35
Daimler, B. 9
Dallam, R. D. 3
Daly, M. M. 3, 27
D'Angelo, 39
Danielli, J. F. 36
Darlington, C. D. 38, 40, 43, 45, 46, 48, 49, 51
Davies, J. T. 210
Davis, L. E. 64, 80
Davison, P. F. 26, 28, 30, 33
Dawson, M. H. 88, 92
De Boer, J. 261
Dedrick, H. M. 88
Dekker, C. A. 44, 46
De Lamater, E. D. 94
Demerec, M. 96
Derjaguin, B. 64
Dijkgraaf, S. 211, 217
Dirr, K. 3
Doesschate, G. ten 221
Doherty, J. K. 40

Donnan, F. G. 58, 61, 63
Drew, R. 105
Duryee, W. R. 40, 42, 47

Edman, P. 10, 14
Elliott, A. 29, 33, 46
Elmes, P. C. 31
Elson, D. 28
Emck, J. 224
Emde, F. 70
Engström, A. 34
Ephrussi-Taylor, H. 91, 95
Eriksson, E. 62
Erlanger, J. 203
Exner, S. 248

Fahmy, O. G. 45
Farrant, J. L. 46
Fatt, P. 122, 128, 139, 141, 145, 152, 153, 155, 158, 159, 160
Felix, K. 3, 5, 16, 26
Fell, H. B. 40
Fischer, H. 3, 5, 6, 7, 26
Fluke, D. 105
Franck, V. F. 200, 201
Fraser, M. J. 31
Fraser, R. D. B. 31, 32
Frazer, S. C. 43
Frisch, K. von 212, 214, 217, 252
Fujii, K. 40

Gall, J. C. 42, 47
Gay, H. 36
Gerstner, H. 180
Goldacre, R. J. 38
Goldschmidt, R. B. 35
Gopal-Ayengar, A. R. 31, 36, 38, 39, 41, 45, 46
Gouy, G. 68
Grahame, D. C. 64
Graybiel, A. 223
Grey Walter, W. 184, 185
Griffen, A. B. 39
Griffith, F. 87, 88, 90
Groen, J. J. 223, 224
Grundfest, H. 134
Guggenheim, E. A. 63, 75

Haddow, A. 51
Hahn, E. 93, 94, 95
Haidinger 252
Hallpike, C. S. 224
Hamer, D. H. 5
Hammarsten, E. 67
Hammarsten, H. 67, 101
Hanby, W. E. 29, 30
Happey, F. 29
Harrison, M. F. 43
Hartley, G. S. 81
Hartline, H. K. 256
Hecht, L. 43
Hecht, S. 256, 257
Helmholtz 252
Hermann, L. 189
Herne, R. 7
Herriott, R. M. 113
Hill, A. V. 180, 183, 193
Hill, S. E. 201
Hill, T. L. 64
Hodgkin, A. L. 123, 125, 134, 179, 191
Hoff-Jørgensen, E. 43
Holst, E. von, 200, 201, 235
Hood, J. D. 224
Hoorweg, P. A. 193
Hotchkiss, R. D. 91, 93, 94, 98, 105, 109
Hückel, E. 64
Hug, O. 17
Huggins, W. H. 228
Hughes, A. F. 40
Hultin, T. 7
Hunt, C. C. 122
Hutter, O. F. 139, 149
Huxley, A. F. 125, 179, 191

Jaeger, L. 36
Jahnke, E. 70
James, D. W. F 26
Jenny, H. 62, 75

Kameyama, N. 63
Katz, B. 124, 127, 128, 134, 135, 139, 140, 141, 142, 145, 147, 148, 152, 153, 155, 158, 159, 160, 161, 162, 163, 166, 167, 179. 191, 203
Kaufmann, B. P. 36
Keynes, R. D. 123, 134
Klaarenbeek, F. W. 64, 67, 70, 71, 72, 74, 79
Köester, H. 11
Koller, P. C. 38, 39, 46, 52
Kossel, A. 2, 3, 4, 10
Kostial, K. 139, 149
Krauss, M. R. 91

VOL
6
195

KREKELS, A. 9, 26
KREUZER, L. 3, 6
KRUYT, H. R. 64
KUFFLER, S. W. 122
KUIPER, J. W. 208, 231, 232, 254
KURODA, Y. 5
KUWADA, V. 48
KUWADA, Y. 40

LA COUR, L. F. 40, 45
LANDAU, L. 64
LANG, H. M. 35
LEDERBERG, E. 96
LEDERBERG, J. 96
LEEUWENHOEK, A. 256
LEIDY, G. 89, 90, 91, 92, 93, 94, 95, 106, 108, 109
LENS, J. 46
LEUCHTENBERGER, C. 29
LEVAN, A. 46, 47
LEVENE, P. A. 97
LILLIE, R. S. 49, 189, 194
LIPPERT, W. 17
LIPSHITZ, R. 28
LIVELY, E. R. 96
LLOYD, D. P. C. 150
LOEB, J. 61
LOEWI, O. 137
LORENTE DE NÓ, R. 203
LOVELESS, A. 33
LOW, P. F. 62
LÖWENSTEIN, O. 218
LUCY, J. A. 28
LUTHER, R. 187

MCCARTY, M. 88, 90, 93, 97, 103, 104, 105, 107, 108, 109, 110, 112
MCDONALD, M. 36
MCDONOUGH, E. S. 36
MACINNES, D. A. 76
MCINTOSH, G. C. 139
MACLEOD, C. M. 88, 90, 91, 93, 97, 103, 104, 105, 107, 108, 109, 110
MALLOCK, A. 249
MANTON, F. 49
MARAZZI, A. S. 203
MARK, J. VAN DER 184
MARMUR, J. 91
MARSHAK, A. 15, 16, 43
MARSHAK, C. 43
MARSHALL, C. E. 62
MATHEW, A. 17
MAURITZEN, C. M. 27
MAZIA, D. 36, 47
MECHELKE, F. 42
MENZER, G. 253, 254, 255
MERCER, E. H. 46
MERCER, H. 38
METZ, C. W. 39, 47
MEUNIER, L. 199, 200, 201
MIESCHER, F. 2, 7, 17
MILLS, G. L. 5
MIRSKY, A. E. 3, 4, 5, 6, 16, 18, 20, 27, 43
MOHN, H. 36
MOORE, S. 27
MOTT, N. F. 261
MURALT, A. VON 191
MYSELS, K. J. 62

NACHMANSOHN, D. 129, 133
NAKAMURA, T. 48
NASTUK, W. L. 139, 152, 158
NAVASHIM, M. 48
NAYLOR, E. J. 253
NEUHAUS, W. 210
NEURATH, H. 201
NEWMAN, E. B. 243
NORMAN, A. 94

O'BRIEN, BRIAN 251
ODATE, Z. 36, 38
OETLINGER, H. 221
OSTER, G. 31
OSTERHAUT, W. J. V. 201
OSTWALD, W. 189
OTA, Y. 36, 38
OVERBEEK, J. TH. G. 62, 64, 68, 73, 75, 78
OVERSTREET, R. 62, 75

PAINTER, T. S. 39
PALAY, S. L. 38
PALLMANN, H. 62
PATTERSON, E. K. 35
PAULING, L. 29, 30
PEECH, M. 62
PERRY, W. L. M. 165
PERUTZ, M. F. 29
PFEIFFER, H. H. 36
PIRENNE, M. H. 245, 256, 259
POL, B. VAN DER, 184
POLLARD, E. 105
POLLISTER, A. W. 4, 18, 29, 43
POUYET, J. 30
PRIGOGINE, I. 175, 178, 186

Proetz, A. W. 209
Pumphrey, R. J. 214, 230

Randall, J. T. 21, 31, 34, 46
Rasch, E. M. 39
Rasmussen, E. 7
Ravin, A. W. 93, 95
Redman, W. 89, 90
Rees, A. L. 46
Riker, W. F. 122
Riley, D. P. 31
Ris, H. 3, 20, 40, 42, 43, 48
Robertson, J. D. 138, 164, 166
Rose, A. 257, 258
Rosenblith, W. A. 226
Rosenblueth, A. 122
Rowan, R. 36
Rowen, J. W. 94
Ruch, F. 34, 37
Rushton, W. A. H. 255, 257
Rutishauser, A. 45

Sadron, C. 30
Samis, C. S. 81
Sanger, F. 10
Schachman, H. K. 44, 46
Schmaltz, G. 237
Schmidt, G. 43
Schmidt, O. H. 203
Schmidt, W. J. 34, 36, 254
Schneider, H. 217
Schultz, J. 35, 36
Sevag, M. G. 103
Shooter, K. V. 26
Sia, R. H. P. 88, 92
Sinke, N. 40
Sivian, L. J. 239, 240
Smiles, J. 49
Sollner, K. 62
Stamm, W. 8
Stanworth, A. 253
Stedman, E. 27, 28
Steffensen, D. 47
Steiner, R. F. 47
Steinhausen, W. 221
Stern, K. G. 101
Steudel, H. 17
Stiles, W. S. 251
Stockhammer, K. 253, 254, 255
Street, A. 29
Strohmeyer, S. 40
Stuiver, M. 209, 210
Stumpf, H. 253
Swift, H. 39, 43

Taylor, F. H. 210
Taylor, H. E. 91, 95
Taylor, P. B. 81
Thannhauser, S. J. 43
Thesleff, S. 165
Thomas, L. E. 3
Thomson, R. Y. 43
Tjio, J. H. 46, 47
Tomlin, S. G. 42, 47
Toraldo di Franeia, G. 251
Tristam, G. R. 5
Tröger, J. 229
Trotter, I. F. 29
Tunevall, G. 89

Van't Hoff, J. H. 186
Velden, H. A. v. d. 256, 259
Velick, S. F. 5
Vendrely, C. 20, 43
Vendrely, P. 3
Vendrely, R. 3, 43
Verwey, E. J. W. 64, 68
Verworn, G. 187
Vetter, K. 191
Vischer, E. 28
Vilstrup, T. 235, 236
Vincent, W. S. 28
Vogel, H. J. 15
Voh, R. 9
Vollheim, G. 203
Vries, Hl. de 215, 230, 231, 233, 234, 235, 236, 237, 240, 241, 243, 244, 247, 256, 257, 258, 260, 262
Vrolijk, J. 236

Wald, G. 258
Waldschmidt-Leitz, E. 3, 9, 11, 12
Waley, S. S. 29
Walker, J. M. B. 43
Watson, J. 29
Watson, J. D. 16, 28, 30, 31, 44, 45, 49
Waugh, D. F. 46
Weiss, G. 193
Wessely, F. 10
Wetzler, G. 189
Wever, E. G. 229
White, J. C. 31
White, S. D. 239, 240

WIEGNER, G. 62
WILKINS, M. H. F. 21, 31, 34
WOODS, H. J. 29
WULF, B. 221
WYATT, G. R. 15, 16

YAMAGIWA, K. 203, 204
YASUZUMI, G. 36, 38
YATES, N. B. 43

ZAMENHOF, S. 16, 28, 90, 98, 105, 106, 108, 109, 114
ZEUTHEN, E. 43
ZINDER, N. D. 96
ZWAARDEMAKER, H. 209, 210
ZWISLOCKY, J. 228, 229

VOL.
6
956

SUBJECT INDEX

ACCOMODATION phenomena 173, 183, 192
Acetylcholine 127–36
— inactivation of 156, 165
— intra-cellular 144
— methods of application 151
— potentials 132, 135
— receptor action 150, 167
— release 137, 144, 146, 166
— — quantization of 138–45
— — role of inorganic ions 139
— replenishment 149
Acipenser sturio 5, 15, 19
Adaptation in vision 257
Adenine 15
— -guanine ratio in NA 28
After potential 190
All-or-nothing principle 172, 186, 203
Allelic series 91
Antibiotics, resistance to 91
Anticholinesterases 128, 129, 133, 142, 143
Arc behaviour 180, 183
Arginine 27
Autocatalytic reaction 180, 186
Axon 122, 203
— electrical transmission in 129
— impulse 137
— membrane 129, 133, 149, 179
— non-medullated 130
— terminals 131
Axoplasm conductivity 125

BACTERIOPHAGE 96
Balbianni ring structure 42
Birefringent fibres 33
Blair's law 193
Blocking agents 166
Brownian motion 236, 237
— — and threshold of hearing 239

CALCIUM ions 32, 47
— — and ACh release 140, 145, 146, 150
Carbaminoylcholine 136
Carboxypeptidase 12
Carcinogenic agents 87, 114
Carcinostatic agents 87, 114
Cell membrane, instability of 172
Charge density, on particles 70
— — surface 72
Cheyne-Stokes respiration 202
Cholinesterase 129, 164
— inhibitor 127, 131, 133, 151
Chromocentre 35
Chromosomes
— behaviour in meiosis 51
— lamp brush 42
— multiple stranded structure 45
— particulate structure 46
— salivary gland 34–9
— spiralization 48, 49
Clupea harengus 5, 6, 15, 18
— *pallasii* 6, 10
Clupeine 5–7
— amino acid sequence 12
— benzoyl-acetyl- 11
— components of 14
— enzyme splitting of 12
— examination of fractions 8, 9, 13
— methylester hydrochloride 7
— peptides in 12, 13
— proline and alanine from 10, 14
— terminal amino acid 9
Cochlea 212, 227, 229
Co-ions 60
— fraction expelled 69, 74
Conchae 208
Conductances of fibre membrane, *see* Permeability
Congo red, Donnan e.m.f. 80
Counterions 60
— mobility of 78
— suspension effect 81
Cupula 210, 213, 219
— "abnormal" motion 223
— piezo-electric effect 231
Cupulometry 222
Curare 127, 142, 155, 166, 204
Cytosine 15

DECREMENT-FREE propagation 173
Denervated muscles 151
Deoxyribonuclease 89, 92, 97, 98, 109
Deoxyribonucleic acid 16, 86, 88
— — and transforming principle 97
— — denaturation 103
— — electrophoretic mobility 105

VOL 6 195

Deoxyribonucleic acid, fractionation of 28, 102
— — from fish sperm 14, 19, 20
— — "functionally intact" 87, 102
— — heterogeneity 101
— — in *H. influenzae* 106
— — "native state" 87, 102
— — paracrystalline form 31
— — purines and pyrimidines in 15
— — twin spiral model 28, 30, 44
— — ultracentrifugation 20, 104
Depolarization 123, 126, 135, 139, 142, 151, 153
Depression, neuro-muscular 148
Diffraction, effects in vision 246
Diffusion 132
— barriers 129, 133
— inactivation by 165
Directional sensitivity in vision 255
Donnan e.m.f. 75, 79, 80
— equilibrium, experimental data 71
— — non-ideal systems 58
— potential 74
— — indirect determination 61
— — suspension effect 62
Dynamic systems, open 173
— — — forces and fluxes in 180
— — — steady states in 173, 181
— — stability 177
— — unstable 172, 173, 178, 189
Dynatron behaviour 180, 183
— -type electrode 188

EAR 210, 226
— inner 242
— middle 228, 229
Electric organs of electric fishes 198
Electrochemical oscillators 199
Electroencephalograms 202, 203
Electrotonic potential 134, 135, 153
— spread 126, 129, 130, 132
End-plate membrane 150, 152, 153, 163
— —, motor 127, 129, 131
— — potential 126, 132, 137, 141, 147, 151, 154
— — — compared with action potential 157, 161
— — — miniature and full 143, 146
— — region 122, 127
Epilepsy 203
Escherichia coli 89, 93
Eserine 165
Excitability 172, 183
Excitable membrane 180
Excitation process 183, 189
— propagation 194
Eye, compound 249
— vertebrate 248

FACILITATION, neuro-muscular 146, 148
Fenton's reagent 112
Fontinine 5, 10, 12
Formaldehyde 113
Franck-Condon principle 260
Free radicals 112
Frequency discrimination 228

GALLINE 3, 5, 6
Gallus domesticus 5
Guanine 15

HAIDINGER effect 252
Haircell 210, 231, 232, 242–4
— model of 234
Hearing 226
Hemophilus influenzae 89–94, 105–12
Heredity determinants 86, 92
Heterochromatin 35, 40
Hill's theory 183, 193
Histones, fractionation of 26
Hydrazides, precipitation as dibenzale 11
Hydrogen bonds 29, 46

IMPEDANCE, acoustic 226–8
Infra-red studies of N.A. and N.P. 31
Interneural interaction models 203
Ion distribution 59, 64, 67
— penetration 123
Ionic concentration gradients 126
Iridine 5, 9–12
Irritation phenomena 172

LABYRINTH 210, 212, 215, 217
Lacustrine 5
Lagena 215
Lateral line 210, 211
— — microphonics of 230, 231
— — organ input-output relation 233
Lipoid membrane 209
Liquid junction potentials and Donnan e.m.f. 76

Local anaesthetics, action of 126
— circuit action 126, 128, 129, 136
— currents 189, 191, 192, 201
Luciani periods 202
Lysine 27

MACULAR pigment 252
Magnesium ions and ACh release 140, 145, 146, 150
Methyl cytosine 15
Micro-electrode technique 127
Microphonic effect 229, 233
Motor nerve 127, 128, 144
Muscle fibres 123, 155
Mustards 113
Mutagenic agents 87, 111, 114

Neisseria meningitidus 89, 90, 93
Nerve fibres 123
— — electric excitation in 123, 126
— endings 122, 131, 139, 143
— membrane 179
— models 172
— — chemical 184
— — electrical 183
— — electrochemical 187, 199
— — hydraulic 174, 180, 183
— — passive iron 189
— — special 179, 193
— -muscle junction 126, 130, 138
— — — chain of events at 137, 147
— signals, propagation 126
Nerves, myelinated 194, 196, 197
— resting state 189
Neural net 198, 199
Nicotine 127
Nose 208
— model of lumen 209
Nuclei and hereditary characteristics 21
— electron micrographs of 17
— from sperm heads 4, 14
— — trout and herring 17
— N, P and arginine content 18
Nucleolus, RNA in 29
Nucleoprotamines extraction 14
— from salmon and sturgeon 19
— genetic aspects 19
— N, P and arginine content 18
— ratio DNA to protamine 20
— size 21
Nucleoprotein 32, 33

OLFACTORY stimulation 210
Onsager reciprocal relations 175, 176
Osmotic pressure 60, 66
Ostwald model 189–96, 203, 204
Otolith 211
— mechanics of 213, 214
— sensitivity 225
— weight 215

PENICILLIN-RESISTANCE 91
Permeabilities of fibre membrane 123
— changes 125, 152
— potassium 125, 167
— sodium 123, 125, 167
*p*H effect on transforming principle 108
— of a suspension 81
Phenylthiohydantoin 10
Pneumococcus 87, 88, 90, 93, 105, 107
Poisson-Boltzmann equation 68
Poisson-grouping and e.p.p. response 146
Polarized light, detection of 252
— — experimental techniques 254
— — visual dichroism 31
Poly-galactophosphate 90
Poly-glucophosphate 90
Poly-ribophosphate 90
Post-synaptic excitation 204
— — inhibitors 151
— — membrane 128, 131, 151, 160, 164
Potassium ions and ACh release 139, 153
Potential differences between two phases 60
— — in fibres 123 (*see also* End-plate potential)
— action or spike 123, 125, 149, 155-61, 191
— active membrane 123
— fluctuations of membrane 143, 162
— resting 123, 160
— threshold 123, 124, 126
— unstable 191
Pre-synaptic action 204
— — membrane 138, 144, 150
— — terminals 128, 131
Procaine 142
Prostigmine 128, 151, 204
Protaminase 11
Protamine 2, 19

VOL
6
195

Protamine amino acid content 4
— ester hydrochloride 4
— from salmonidae 10
— hydrazinolysis 11
— inhomogeneity 7
— molecular weight 20
— mono-, di- and tri- 4
— preparation 3
— sulphate 3
— terminal amino acids 9–12
Protein, β 31
— chain, hydrogen bonds 29
— systems, Donnan e.m.f. 80
Protoplasm membrane 190
Proximal centriol 18
Pulsating ganglion cells 201

RANVIER nodes 194, 195, 197
Receptor molecules 152
Refractory behaviour 173, 183, 192
— time 191
Relaxation oscillations 184, 200
Resting nucleus 40, 41
— — DNA content 43
Rheobase 182, 189, 191, 197
Ribonucleic acid in nucleolus 29
Rythmical behaviour 173, 184, 199–203
— stimuli 203

SACCULUS 214
Salmine 5, 6, 10, 12
— hydrochloride separation 9
Salmo fontinalis 5, 6, 12, 15, 18
— *irideus* 5, 9, 15, 18
— *lacustris* 5
— *salar* 5, 6, 9, 15, 19
— *trutta* 5, 6, 12, 15, 18
Saltatory propagation 194–8
Scomber scombrus 5
Scombrine 5
Semicircular canal 211, 212, 235
— — anatomical data 221
— — mechanics of 219
Short-circuit theory 153, 161
Skeletal muscle fibre 122
Smell 208
Sodium arabinate Donnan equilibrium 71–4
— — — e.m.f. 79
Sodium ions and ACh release 139, 153
Sperm heads 34
Spermatozoa, rooster 3, 19
Spheroides rubripes 5
Spheroidine 5
Stiles-Crawford effect 251
Stimulus 181
— cathodic 195, 199
Streptomycin-resistance 91, 107
— -sensitivity 91
Strömchentheorie 190
Sturine 5, 10, 12
Sub-threshold pulse 126, 130, 155
— stimuli 181, 182, 191, 203
Supra-threshold stimuli 181, 182
Swimming bladder 216
Synapse 130–2, 150
— artificial 204
— chemical transmission at 129
Synaptic connections 203
— transmission 122, 130, 204
Synchronized oscillations 200, 202

TASTE 208
Temperature, effect on sensitivity of eye 262
Testes of spawning fish 3
Theory of solutions 64
Thermal agitation in vision 258
— explosions 187
Threshold for stimulation 172, 173, 181, 182, 191, 203
Thymine 15
— -cytosine ratio in nucleic acids 28
Time of indication 214, 215, 222, 225
— — recovery 225
Transduction phenomena 96
Transformation 86–92
— bacterial 92
— competition 94
— susceptibility 92, 93
Transforming phenomenon 86–92
— principle 86–96
— — chemical nature 97
— — effects of various agents 107–12
— — isolation and purification 103
— — unstable 114
— — viscosity 99, 109, 110
Truttine 5, 10, 12
Trypsin 12

Ultra-violet studies of nucleic acid 31
— — irradiation and transforming principle 112

Utriculus 214

Vision 245
— and lightquanta 256
Visual acuity 249

Waveguides analogy of sense cells 250
Weber's bones 217
Weiss's Law 193
Wenkebach periods 202

VOL
6
195

1

PROTAMINES AND NUCLEOPROTAMINES

K. Felix, H. Fischer and A. Krekels

CONTENTS

		Page
I.	DISCOVERY AND OCCURRENCE	2
II.	PREPARATION OF PROTAMINES	3
III.	AMINO ACID CONTENT OF PROTAMINES	4
IV.	INHOMOGENEITY OF THE PROTAMINES	7
V.	TERMINAL AMINO ACIDS	9
VI.	THE AMINO ACID SEQUENCE IN CLUPEINE	12
VII.	DEOXYRIBONUCLEIC ACID FROM FISH SPERM	14
VIII.	NUCLEOPROTAMINES	17
IX.	GENETIC ASPECTS	19
	REFERENCES	21

VOL.
6
956

1

PROTAMINES AND NUCLEOPROTAMINES

K. Felix, H. Fischer and A. Krekels

Discovery and Occurrence

The first protamine was isolated 1874 by Friedrich Miescher[1] in Basel from sperm heads of Rhine salmon. He extracted the defatted sperm with diluted hydrochloric acid, neutralized the extract and precipitated the whole material with platinum chloride. Due to the high nitrogen content he believed it to be a simple organic base and called it protamine. The hydrochloric extract contained traces of decomposition products of nucleic acid. Protamine is not present in the immature testes; it appears just before maturation. Miescher proved its saltlike combination with nucleic acid but could not explain its significance.

The true nature of the base was recognized by Albrecht Kossel[2] 20 years later. He repeated Miescher's experiments with salmon spermatozoa and isolated similar bases from the sperm of sturgeon, several trout species and numerous other fish. He proved that these bases are composed entirely of amino acids and, therefore, are proteins; they contain only a few, chiefly basic, amino acids.

Kossel found them of interest in both chemical and biological respects. Because of their simple composition, they were the most suitable materials to study the fundamentals of protein structure and the manner of amino acid linkage. His biological interest was not only based on the fact that they probably have a part in heredity; they also enabled him to realize one of his favourite ideas: He did not consider a substance occurring in a cell, regularly or temporarily, a single isolated fact, but a link in a series of alterations in the course of living processes. He realized that the protamines are a turning-point in the transformation of protein, since they are formed by continuous simplification from complicated proteins and are then again converted into complicated proteins during the development of the fish from the fertilized egg cell.

The fish facilitate the study of this process for the biochemist, because they do not take in food during sperm maturation. The protein of the generating organs is, therefore, built solely from body protein.

Miescher already showed that muscle protein must be the source of protamine, since the body muscles of the salmon shrink to a minimum when they ascend the Rhine and the testes are enlarged 50-fold.

The protamines were first found only in the sperm of the bony fish (Teleostei); a finding which was extended in recent years. DALY, MIRSKY and RIS[3] isolated a basic protein from rooster sperm, galline. FISCHER and KREUZER[4] repeated their experiments and confirmed that galline is a real protamine. It is not unlikely that basic proteins also occur in the spermatozoa of higher animals. VENDRELY and VENDRELY[5] have observed that the cell nucleus of ram sperm is so rich in arginine that the ratio arginine/phosphorus almost suggests the presence of nucleoprotamine. This seems to be true for other mammals too and perhaps even for man, for DALLAM and THOMAS[6] have found an abundant amount of basic amino acids in the nucleus of these spermatozoa. However, until now it has not been possible to extract a protamine from mammal spermatozoa with acids.

PREPARATION OF PROTAMINES

It has already been emphasized that the protamines are only found in mature spermatozoa. In order to obtain a pure preparation, the testes of spawning fish, or even better, fresh fish milt must be used. The latter has the great advantage of containing no other cells besides spermatozoa. KOSSEL[2] minced the testes, stirred them with a large amount of water and then filtered them through several layers of muslin. The milky filtrate was acidified with dilute acetic acid, thereby causing separation and dissolution of the tails and precipitation of the sperm heads. The latter were isolated by filtration or centrifugation, washed with water and dried with alcohol and ether.

He extracted the moist or already dry sperm heads several times with 1 per cent sulphuric acid and from this extract precipitated the protamine sulphate with alcohol. This was dissolved in a small amount of hot water and reprecipitated by cooling as a yellowish oil. It was again dissolved in water and precipitated with sodium picrate. The protamine picrate was taken up by sulphuric acid and extracted with toluene. The protamine sulphate was then precipitated in alcohol as a pure white powder.

In the course of time the original method was altered somewhat in order to prevent the decomposition of the nucleic acid and of the protamine during preparation.

WALDSCHMIDT-LEITZ and co-workers[7] have developed their own purification procedure based on the repeated precipitation of protamine sulphate from hot water.

FELIX and DIRR[8], [9] have decomposed protamine picrate more cautiously with hydrochloric waterfree methyl alcohol and thereby obtained the methyl ester hydrochloride of the protamine. They first suspended the dried picrate in absolute methanol and then saturated the mixture with gaseous hydrochloric acid under complete exclusion

of atmospheric moisture. The procedure was simplified later by suspending the picrate in methanol, containing 0·2 N hydrochloric acid for 24 hours at room temperature. The protamine ester hydrochloride is precipitated with ether as a pure white powder.[10]

BLOCK and BOLLING[11] prepared the protamine in a different manner: Their procedure is supposed to give a much greater yield but it is very probable that it causes some hydrolysis. The sperm heads are precipitated by 1 per cent citric acid, extracted with 0·2 per cent hydrochloric acid and the extract heated to boiling. The *p*H is adjusted to 7–8 with ammonia and the insoluble material centrifuged off. The protamine is precipitated as an oil from the filtrate with the optimal amount of 33 per cent metaphosphoric acid and the protamine metaphosphate converted into the sulphate by boiling with N sulphuric acid; this is precipitated and dried by the addition of alcohol or aceton.

All methods described up to now use as starting material sperm heads which have been precipitated with diluted acetic or citric acid. However, the heads do not only consist of the cell nuclei but also of cytoplasm which surrounds them as a small border. Besides, the acids precipitate cytoplasmic components which can only be eliminated by a very difficult purification process. For this reason it has proven much more advantageous to use nuclei which are free of cytoplasmic material.

The naked nuclei may be procured by plasmolysis as is described below. From them the protamine is extracted with 0·2 per cent hydrochloric acid and the protamine hydrochloride precipitated with acetone.[12] It is also possible to dissolve the nuclei in a 10 per cent sodium chloride solution, according to MIRSKY and POLLISTER,[13] and to salt out the protamine with sodium chloride in form of oil.

AMINO ACID CONTENT OF PROTAMINES

As already mentioned, the protamines contain only a few amino acids, viz. nine. Arginine, alanine and serine were found in all protamines, proline and valine in most of them, glycine and isoleucine in many and the rest (histidine, lysine, threonine, aspartic and glutamic acid) only in few. Tryptophane, phenylalanine and the sulphur-containing amino acids have never been found in the protamines. Until now tyrosine was found only by KOSSEL in thynnine and cyclopterine. In the latter the reaction for tryptophane was also positive. Maybe these preparations were obtained from testes which were not entirely mature.

Among these amino acids the basic ones dominate by far. The ratio between them and the monoamino acids is about 2 : 1. The monoprotamines contain only a single basic amino acid, i.e. arginine. The others (diprotamines) contain lysine or histidine besides arginine and in the triprotamines all three basic amino acids are present.

VOL
6
195

956 VOL.

TABLE 1

Amino Acid Composition of some Protamines

Molecular ratio of the different amino acids

(The amino acid occurring in the lowest concentration, usually isoleucine, is equal to one molecule)

Protamine	*Author*	*Species*	*Arginine*	*Lysine*	*Histidine*	*Glycine*	*Alanine*	*Serine*	*Threonine*	*Proline*	*Valine*	*Isoleucine*	*Aspartic acid*	*Glutamic acid*	*Tyrosin*	*Arginine-Nitrogen expressed in % of the total nitrogen*
Clupeine	Block[11]	*Clupeus harengus*	65	–	–	–	7	4	2	9	4	1	–	–	–	88·5
	Felix[14]		53	–	–	–	4	6	2	5	4	1	–	–	–	89·7
	Kuroda[15]		60	–	–	–	5	1	1	4	1	1	–	–	–	94·0
Salmine	Block[11]	*Salmo salar*	51	–	–	–	4	6	–	7	4	1	–	–	–	90·4
	Felix[16]		55	–	–	3	2	5	–	5	4	–	–	–	–	93·0
	Tristam[17]		41	–	–	3	1	7	–	4	2	1	–	–	–	89·0
	Hamer[18]		75	–	–	7	1	8	–	6	5	2	–	–	–	91·0
	Velick[19]		34	–	–	3	1	4	–	4	3	?	–	–	–	86·3
	Corfield[20]		50	–	–	4	1	7	–	6	3	1	–	–	–	89·8
	Mills[21]		78	–	–	4	1	4	–	3	1	1	–	–	–	90·6
Scombrine	Kuroda[15]	*Scomber scombrus*	112	–	–	–	9	4	2	7	3	1	–	–	–	93·6
Spheroidine	Kuroda[15]	*Spheroides rubripes*	71	–	–	–	9	7	1	1	1	1	–	–	–	93·0
Iridine	Felix[12]	*Salmo irideus*	50	2	–	2	2	3	–	5	4	1	–	–	–	91·4
Truttine	Felix[22]	*Salmo trutta*	50	–	–	2	2	3	–	5	5	–	–	–	–	89·5
Fontinine	Felix[22]	*Salmo fontinalis*	50	–	–	2	2	3	–	5	5	–	–	–	–	90·4
Lacustrine	Felix[12]	*Salmo lacustris*	50	–	5–15	2	2	2	–	5	2	1	1	1	–	
Sturine	Felix[16]	*Acipenser sturio*	35	9	7	2	5	3	1	–	–	2	–	1	–	67·0
Galline	Mirsky[3]	*Gallus domesticus*	75	–	2	10	2	14	3	7	2	1	1	2	6	76·3
	Fischer[4]		44	–	–	1	5	5	2	5	3	1	–	1	–	88·0

TABLE 2

Basic Proteins of Lower Marine Animals.
Relation of Amino Acids

Species	*arg*	*lys*	*hist*	*glyc*	*ala*	*ser*	*threo*	*prol*	*val*	*leu*	*ileu*
Patella vulgata .	40	32	1	15	20	15	7	15	7	6	2
Patella coerulea .	80	14	–	14	9	21	3	21	6	6	1
Arbacia lixula .	14	30	1	22	30	9	3	12	3	5	4

The results obtained after the development of quantitative methods of ion exchange and paper chromatography are summarized in Table 1.

The figures in column 1–12 give the molecular ratios of the amino acids. The calculation is based on the assumption that the lowest concentration of an amino acid occurring in a protamine is equal to one molecule. However, since the experimental errors and the differences in the applied methods become especially evident by this calculation, we have depicted the results of the arginine determinations in percentages of total nitrogen in the last column. Here the results of the various authors show better agreement. There is no discrepancy concerning the kind of amino acids occurring in the individual protamines.

Clupeine and salmine have been investigated most extensively and according to our present knowledge, it seems established that clupeine (from *clupea harengus*) contains threonine but no glycine, salmine (from *salmo salar*) on the other hand has glycine and no threonine.

ANDO and co-workers[23], [24] investigated clupeine obtained from *clupea pallasii* and found that it differs from the above mentioned type by its glycine content.

It must be emphasized once more that the protamines investigated recently are entirely free of aromatic and sulphur-containing amino acids. According to FISCHER and KREUZER,[4] galline makes no exception. The tyrosine MIRSKY[3] found may be due to an impurity; his low figures for arginine content may possibly be interpreted in the same way.

Although the results of the quantitative amino acid analyses are not quite satisfying at present, it may be assumed that the amino acid composition of the protamines is characteristic for the individual fish species. Of all the protamines mentioned in the table only two have the same composition: those of the two trout species *salmo trutta* and *salmo fontinalis*. Perhaps it is not merely coincidental that the individuals of these two species can be crossed. The hybrids, however, are sterile.

In this connection it is, moreover, necessary to mention substances which have been isolated by HULTIN and HERNE[25] from the nucleus of spermatozoa of *patella vulgata*, *patella coerulea* and *arbacia lixula*. These substances also have basic character and differ from the real protamines by their higher molecular weight and the larger monoamino acid content. In a way they form an intermediate stage between protamines and histones (see Table 2).

INHOMOGENEITY OF THE PROTAMINES

While MIESCHER believed salmine to be a simple base, we know for about 25 years that protamines, especially clupeine, are not homogeneous substances but a mixture of several very similar components. These differ chiefly by their molecular weight and their monoamino acid content. Nevertheless, in all of them the ratio of basic to monoamino acids is the same as in the original mixture, i.e. about 2 : 1.

Our earlier experiments[8] already showed, that the methyl ester hydrochloride which results from decomposition of clupeine picrate with methanolic hydrochloric acid, can be separated in fractions which differ by their solubility in concentrated methanolic hydrochloric acid. The fractions having a smaller molecular weight (calculated by the methoxyl content) were more readily soluble than those having a larger one. RASMUSSEN[26] objected to these conclusions. He suspected that clupeine had been hydrolyzed during the decomposition of the picrate. However, recent investigations have confirmed the inhomogeneity.

For the majority of experiments we used clupeine which was prepared with great care and contained 24·1 per cent nitrogen. 89–90 per cent of its total nitrogen are arginine nitrogen and 10–11 per cent are monoamino acid nitrogen. The molecular ratio of arginine to monoamino acids varies between 2·23 : 1 and 2·25 : 1. Before hydrolyzing, 1·36 per cent of the total nitrogen were titratable with formol. After hydrolysis, the free amino nitrogen increased to 33·43 per cent.

The preparations were first examined by electrophoresis. Paper electrophoresis revealed only two fractions.[10]

Considerably more fractions were found when the TISELIUS method was used. The results obtained by this method were greatly influenced by the ionic strength of the buffer solutions.[10] Clupeine methyl ester hydrochloride was dissolved in the veronal sodium acetate hydrochloric acid buffer of MICHEALIS. Within the range of *p*H 3·8 and 8·9 the protamine migrated to the cathode. According to recent unpublished investigations (FISCHER), it also migrates to the cathode in a glycine sodium hydroxide buffer of *p*H 12.

The ionic strength of the buffer was varied by adding sodium chloride. In low ionic strength ($\mu = 0{\cdot}3$) the curve showed only a few details. These became more evident at $\mu = 0{\cdot}46$ and $\mu = 0{\cdot}98$.

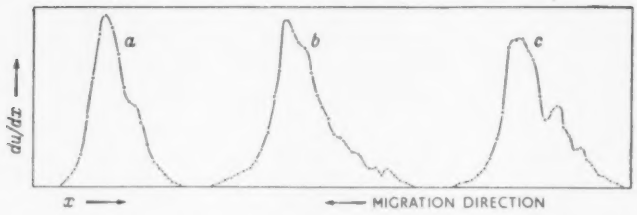


Fig. 1. Electrophoresis of clupeine methylester hydrochloride in sodium barbiturate acetate buffer *p*H 3·9 of various ionic strengths.
(*a*) $\mu = 0,30$; (*b*) $\mu = 0,46$; (*c*) $\mu = 0,98$.

Counter current distribution gave preciser results. A distribution over only 24 units[27] produced a curve with numerous irregularities, especially in the descending limb. In later experiments which have not yet been published, STAMM[28] distributed clupeine methyl ester hydrochloride over 400 units and received the following curve.

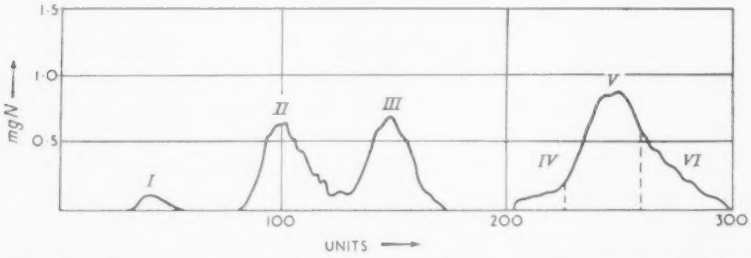


Fig. 2. Counter current distribution of clupeine methyl ester hydrochloride over 400 units in a system of 15 per cent sodium acetate as aqueous phase and 5 per cent lauric acid in n-butanol as organic phase.

Four peaks can be distinguished in this curve. The first three probably correspond to three main components, not yet quite pure. The last peak shows a number of irregularities which represent at least three components. At present, we assume that clupeine consists of at least six components, but it is very likely that there are even more.

The fractions which were procured by counter current distribution differed somewhat in their composition and molecular weight. All of them contained arginine, alanine, serine, threonine and proline. Isoleucine was present only in fractions I and II. In this assay fraction I was contaminated with traces of decomposition products of nucleic acid. Fractions I to III are composed of higher peptides which dialyze very slowly through cellophane. Fractions IV to VI pass through the membrane much faster and, therefore, contain smaller peptides.

This conclusion was confirmed by the examination of the fractions in ultraviolet light of a wavelength of 220 mμ. Although this wavelength

does not give the maximum absorption of the peptide group, the differences in absorption were significant. Referred to nitrogen, fractions I to III absorb 14 times more light than fractions IV to VI.

According to the results obtained by ultracentrifugation, the number of components present could be about ten. However, this method did not give clear results. (DAIMLER.[29])

In solubility experiments, clupeine also behaved like an inhomogeneous substance. We used a 27·5 per cent solution of sodium chloride in which clupeine is only slightly soluble. When increasing amounts of clupeine hydrochloride are added to the solution, the dissolved portion first increases in a straight line. However, on reaching the point of saturation, there was not a sharp break in the curve, but a smooth bend.[30]

Recently KREKELS[31] succeeded in separating clupeine and salmine hydrochloride into three distinct fractions by paper chromatography. The ascending chromatograms were developed in a mixture of 18 ccm 86 per cent ethanol, 6 ccm 20 per cent trichloroacetic acid and 1 ccm secundary butanol. One of her chromatograms is reproduced in this article (Fig. 3).

From all these experiments we must deduct that clupeine is a mixture of several components, of very similar composition. All of them contain large amounts of arginine. Therefore, it is very difficult to separate them from each other.

The same must be true for all the other protamines, although we have proven this only for two commercial preparations of iridine (*salmo irideus*) and salmine (*salmo salar*). Both behaved like inhomogeneous substances in counter current distribution.[27]

The inhomogeneity of clupeine is also confirmed in the work of WALDSCHMIDT-LEITZ and VOH.[32] They succeeded in decomposing clupeine picrate into an isoleucine-free and an isoleucine-containing portion by dissolving the picrate in aceton water and adding either water or alcohol.

TERMINAL AMINO ACIDS

Since protamines are mixtures of several components, one would expect these components to differ in their terminal amino acids. However, this is only true for the carboxylic end of the peptide chain. At the nitrogen end the same amino acid is always found.

The first clue in identifying the amino acid at the nitrogen end of clupeine was given by the formol titration. With this method a very small amount of amino nitrogen was found.[26], [9] However, since nitrogen was not evolved by nitrous acid and on the other hand 80 per cent of the imino nitrogen of proline is detectable by formol titration, it was very probable that proline must be the terminal amino acid at the nitrogen end of clupeine.

In recent years several authors have confirmed this with the dinitrofluorbenzene method of SANGER.[33] They all agree that proline is the only N-terminal amino acid.[34], [35], [36] The same is true for salmine[37], [19], [36] and for the trout protamines known until now.[36]

In clupeine from *clupea pallasii*, ANDO and co-workers[23], [24] found alanine to be the N-terminal amino acid.

For sturine two N-terminal amino acids were determined by the SANGER method, i.e. alanine and glutamic acid.[36] We must admit that this sturine was prepared from old sperm heads which KOSSEL gathered 40 years ago; it is uncertain whether these heads were mature or not.

Another method for the determination of N-terminal amino acids was developed by EDMAN.[38] Here the peptide in question is treated with phenyl-isothiocyanate at *p*H 9 and 40°C; hereby the hydrogen of the free amino or imino groups is substituted by the phenylthiocarbamyl residue. When hydrochloric acid in waterfree nitromethane acts on this peptide derivative, the terminal amino acid is split off as phenylthiohydantoin. Hydrolysis of the hydantoin yields the free amino acid which may be identified by paper chromatography. The remaining peptide, shortened by one amino acid should be able to undergo the same reactions and it should be possible to split off one amino acid after the other. In this way EDMAN has decomposed di-, tri- and tetrapeptides.

We have treated the protamines according to this prescription but they show a different behaviour. Instead of one always two amino acids were split off.[39]

By this procedure, proline and valine were set free from the *salmonidae* protamines (iridine, truttine, fontinine, salmine); proline and alanine from clupeine, alanine and glutamic acid from sturine. The results obtained with the EDMAN method agree with those of the SANGER method only in the case of sturine.

We believe that in the case of the five first mentioned protamines, the first two amino acids at the nitrogen end of the peptide chain always react. With a different method we were able to determine that alanine is the second amino acid in clupeine (see below).

Earlier investigations of ABDERHALDEN and BROCKMANN[40] as well as WESSELY and co-workers[41] also agree with this conception. The first two authors investigated the reaction of phenylisothiocyanate on glycyl-D, L-leucyl-benzylamine, and found that besides the hydantoin of glycine that of leucine is also split off. They concluded that the imino group of the peptide linkage may have reacted with the phenylisothiocyanate.

WESSELY and co-workers propose another explanation for the splitting off of two amino acids.

Concerning the amino acids at the carboxylic end, we observed that small amounts of urea are set free when arginase acts on clupeine.[42] From this fact we concluded that arginine stands at the C-end of clupeine, since arginase needs only a free carboxyl and a free guanidine group for its action; the amino group may be substituted by an acyl, methyl or a peptide residue or replaced by a hydroxyl group.

The conclusion that arginine stands at the carboxylic end was confirmed by a chemical method.[43] If clupeine is first benzoylated and then acetylated, the guanidine groups of all arginine residues are substituted by a benzoyl and an acetyl residue, respectively. Then the arginine residue at the C-end has three acid residues, two at the guanidine group and one at the α-amino group (peptide chain). That means, it must behave like the triacetyl anhydro-arginine of BERGMANN and KOESTER.[44] By acid hydrolysis it must be split into urea and ornithine. We have, indeed, found less arginine in the hydrolysate of benzoyl-acetyl-clupeine and instead we isolated the equivalent amount of ornithine.[43]

At that time WALDSCHMIDT-LEITZ,[7] moreover, found that protaminase splits off arginine from clupeine; he also concluded that the carboxylic end is formed by arginine.

For the determination of the C-terminal amino acid AKABORI and co-workers[45] recommended the cleavage of protein with hydrazine. All amino acids forming a peptide bond with their carboxyl group are converted into hydrazides; only those at the carboxylic end are set free as such. The amino acid hydrazides were precipitated as dibenzale with benzaldehyde, while the free amino acid remained in solution. Under these circumstances free arginine partly reacts with benzaldehyde, forming a compound of poor solubility. Nevertheless, a sufficient amount remains for paper chromatographic identification. In any case, this method also indicated that in clupeine a number of peptide chains had arginine at the carboxylic end. In addition, alanine, valine and proline were found as free amino acids.[39]

We mention these results conditionally, since we do not know whether the hydrazides are precipitated quantitatively by benzaldehyde or if they are possibly hydrolyzed by this procedure.

These investigations are interesting because they show that the various clupeine components may be distinguished by the amino acid at the carboxylic end.

We also applied hydrazinolysis to the trout protamines as well as to salmine and sturine and here too, several amino acids were found at the carboxylic end.[39] Table 3 gives a summary of our entire investigations on terminal amino acids.

In this table the similarities of the amino end and the variations at the carboxylic end become evident. The protamines of salmonidae all

have proline as the first and valine as the second amino acid. Clupeine also begins with proline but this is followed by alanine. Perhaps the similarities of the fish families are manifested by the amino end while the carboxylic end is responsible for the differences. If our findings prove to be correct, the two protamines of the two trout species, *salmo trutta* and *salmo fontinalis*, having the same composition, would differ only by the carboxylic end of their components.

TABLE 3

Terminal and Second Amino Acids of the Protamines

	N-terminal	*Second*	*C-terminal*
Clupeine (*Clupea harengus*)	Proline	Alanine	Arginine, Alanine, Valine, Proline
Iridine (*Salmo irideus*)	Proline	Valine	Arginine, Alanine, Valine
Truttine (*Salmo trutta*)	Proline	Valine	Alanine, Valine, Proline
Fontinine (*Salmo fontinalis*)	Proline	Valine	Alanine, Valine, Serine
Salmine (*Salmo salar*)	Proline	Valine	Arginine, Alanine, Valine, Proline
Sturine (*Acipenser sturio*)	Alanine Glutamic acid	?	Arginine, Alanine, Valine

In recent years WALDSCHMIDT-LEITZ[46] attempted to determine the amino acid sequence at the carboxylic end of clupeine by treating it with carboxypeptidase. He found that arginine, alanine and valine are split off. At the time of these experiments he was still of the opinion that clupeine consists of a single peptide chain ending with arginyl-arginine; the latter being preceded by the amino acids mentioned. However, these results also allow the other interpretation, i.e. that the different amino acids are split off from the various components.

THE AMINO ACID SEQUENCE IN CLUPEINE

Many years ago, we split clupeine partially with enzymes and acids and tried to isolate some of the peptides formed. At that time we isolated from trypsin hydrolysates arginyl-arginine and dipeptides of arginine with alanine, serine and valine and arginyl-hydroxyproline.[47]* Among the products of partial acid hydrolysates, we isolated a tetrapeptide which consisted only of arginine.[48], [49] Indirectly, we could

* We never found hydroxyproline in later preparations.

prove the presence of dipeptides of monoamino acids in these hydrolysates.

We have summarized these earlier results in a preliminary formula. According to this, the peptide chains begin with proline and end with arginyl-arginine, and between them four arginine residues alternate with two monoamino acid residues.[9]

Pro (arg-arg-arg-arg-m-m)$_n$. arg-arg

After the war, we again took up these experiments and first repeated the partial acid hydrolysis. Thirty-five per cent of the whole clupeine nitrogen was isolated in form of the tetrapeptide of arginine.[50] This confirms our first conception that the combination of four arginine residues side by side must occur several times in the peptide chains of all components.

Furthermore, we isolated the following peptides as picrolonates:

Alanyl-arginyl-arginine[50]
Seryl-arginyl-arginine[50]
Arginyl-arginine[50]
Prolyl-valyl-arginyl-arginine[51]
Valyl-alanyl-arginyl-arginine[51]

The following dipeptides were identified by paper chromatography:

Seryl-alanine[51]
Isoleucyl-alanine[51]
Valyl-alanine[51]

From three other peptides it was not possible to determine the exact amount and the sequence of the amino acids. They contained (1) serine, valine, arginine; (2) serine, arginine, alanine; (3) serine, arginine, alanine, valine.[51]

From a partial acid hydrolysate of DNP-clupeine we identified by paper chromatography

DNP-prolyl-alanine[51]
DNP-prolyl-alanyl-arginine[51]
DNP-prolyl-alanyl-(arginyl)$_4$-threonyl-serine[51]
DNP-prolyl-alanyl-(arginyl)$_4$-threonyl-seryl-(arginyl)$_4$-valyl-isoleucine[51]

Three of these peptides contain arginine but the exact amount is not yet determined.

Trypsin split DNP-clupeine in three DNP-peptides. One of them contained all the amino acids of clupeine. Serine and threonine were missing in the two others. Besides these DNP-peptides we found three

ninhydrine positive peptides. The first contained all the amino acids of clupeine, the second all except isoleucine and in the third, there was no valine or isoleucine.[39]

These preliminary results may be interpreted as follows: All components of clupeine begin with proline. The second amino acid also seems to be identical in all of them, since we only found alanine as second amino acid with the method of EDMAN.[38] This is followed by four arginines, since this tetrapeptide occurs very often and in addition we have isolated a tripeptide, DNP-prolyl-alanyl-arginine.[51] The seventh and the eighth amino acid could be, at least in one component, proline and valine; in another, serine and threonine. After these two monoamino acids, we have to assume another arginine tetrapeptide in the larger components. We still believe that dipeptides of monoamino acids alternate with tetrapeptides of arginine. However, somewhere in the peptide chains, it may occur that only one monoamino acid stands between two tetrapeptides of arginine, while this is compensated at another place by three monoamino acids. More monoamino acids together would suggest the existence of hexa- or octopeptides of arginine in the chains, because the relation of two to one between arginine and monoamino acids has to be preserved.

Thus the earlier formula for clupeine must be changed slightly, however, this is still tentative:

$$\text{Pro-ala-arg}_4(\text{m}_2 \cdot \text{arg}_4)_n \cdot \text{m}_2 \cdot \text{arg}_2$$

The particular components would differ chiefly in the length of the chain, in the distribution of the monoamino acids between the polypeptides of arginine and in the amino acids at the C-end. May we recall, in this connection, the results of the counter current distribution.[28] There we found four fractions in which isoleucine was missing and two which contained all the amino acids of clupeine. One of the components with the molecular weight of about 12,000, containing all the amino acids of clupeine, could have the following formula:[39]

$$\text{Pro-ala-arg}_4\text{-pro-val-arg}_4 \cdot \text{ileu-ala-arg}_4\text{-val-ser-arg}_4\text{-ser-prol-arg}_4\text{-ser-threo-arg}_4\text{-ser-ala-arg}_4\text{-prol-ser-arg}_4\text{-ser-val-arg}_4\text{-threo-prol-arg}_4\text{-val-ala-arg}_2$$

DEOXYRIBONUCLEIC ACID FROM FISH SPERM

We already mentioned at the beginning that the protamines in fish sperm are combined with deoxyribonucleic acid. We isolated the latter carefully by dissolving the cell nuclei in 10 per cent sodium chloride solution;[13] sometimes we precipitated the nucleoprotamines by pouring this solution into distilled water and then dissolved the precipitate again in salt solution. From this or the former solution we

VOL 6 195

salted out the protamines by saturation with sodium chloride. After centrifugation, the supernatant was freed from sodium chloride by dialysis and the nucleic acid precipitated by alcohol as a fibrous mass which could be twisted around a glass rod. The preparation obtained by this procedure is highly polymeric and free of protein.[22]

First of all, nitrogen and phosphorus were determined in the various preparations; from these figures the ratio nitrogen to phosphorus was calculated. The preparations were then hydrolyzed with perchloric acid, according to MARSHAK and VOGEL[52] and the bases determined by paper chromatography, using the method of WYATT.[53] The results are summarized in Table 4.[54]

TABLE 4

Composition of Deoxyribonucleic Acids of Sperm Nuclei of various Fish

	Salmo irideus	*Salmo trutta*	*Salmo fontinalis*	*Salmo salar*	*Clupea harengus*	*Acipenser sturio*
%N: . .	14·92	15·46	15·20	14·29	15·29	14·75
%P: . .	8·78	8·90	8·95	8·47	9·15	8·18
N/P: . .	1·69	1·73	1·69	1·69	1·67	1·80
%Bases						
Adenine .	10·65	10·80	10·65	10·15	10·60	10·35
Thymine .	9·20	9·20	9·25	8·90	8·90	8·85
Guanine .	8·90	8·90	8·70	8·80	8·60	8·75
Cytosine .	6·15	6·25	6·25	6·25	6·20	5·80
Methyl-cytosine .	+		+		+	

From this table we have calculated the molecular ratio of the bases and referred them to 100 atoms of phosphorus.

TABLE 5

Distribution of Purines and Pyrimidines in DNA-preparations

(Expressed in moles per 100 *atoms* P)

DNA	*Adenine*	*Thymine*	*Guanine*	*Cytosine*
Salmo irideus .	28	26	21	20
Salmo trutta .	28	26	21	20
Salmo fontinalis .	27	26	20	20
Salmo salar	27	26	21	21
Clupea harengus .	27	24	19	19
Acipenser sturio .	29	27	22	20

For comparison, Table 6 gives the results other authors found.

TABLE 6

Distribution of Purines and Pyrimidines in Deoxypentose Nucleic Acids, expressed in moles per 100 atoms phosphorus

DNA	*Author*	*Adenine*	*Thymine*	*Guanine*	*Cytosine*
Human thymus	ZAMENHOFF[55]	28	28	19	16
Calf thymus	MIRSKY[56]	28	28	24	20
Calf thymus	WYATT[57]	25	26	19	19
Calf thymus	CHARGAFF[58]	29	26	19	16
Calf thymus	MARSHAK[52]	29	29	26	23
Calf thymus	FELIX[54]	29	28	21	20
Calf kidney	MIRSKY[56]	27	29	22	20
Calf kidney	MARSHAK[52]	27	27	22	21
Calf liver	MARSHAK[52]	25	26	23	21
Calf heart	MARSHAK[52]	26	26	21	21
Sheep spleen	MIRSKY[56]	26	27	21	20
Horse spleen	MIRSKY[56]	28	26	22	19
Beef spleen	CHARGAFF[58]	25	24	20	15
Beef spleen	WYATT[57]	26	25	21	20
Bull testes	FELIX[54]	27	28	21	17
Human sperm	ZAMENHOFF[55]	27–29	31	17–18	18
Sea urchin sperm	MIRSKY[56]	27	31	18	18
Trout sperm	MIRSKY[56]	29	27	22	19

Both tables show that the fish sperm DNA fits quite well into the general scheme of animal deoxyribonucleic acids. It is also obvious that the nucleic acids of the various fish species do not differ significantly as far as the building blocks are concerned. Their monotonous composition is quite a contrast to the numerous variations shown by the protamines. We admit, the nucleic acids may differ in the arrangement of the nucleotides in the chain; this would, of course, allow many variations.

Finally, we wish to mention that the ratio of adenine to thymine and guanine to cytosine is almost 1. This corresponds to the conception of WATSON and CRICK[59] concerning the structure of nucleic acids.

TABLE 7

Molecular Ratios of the Purine and Pyrimidine Bases in the DNA of Fish Sperm

	adenine/thymine	*guanine/cytosine*
Salmo irideus	1·07	1·05
Salmo trutta	1·07	1·05
Salmo fontinalis	1·04	1·00
Salmo salar	1·04	1·00
Clupea harengus	1·12	1·00
Acipenser sturio	1·07	1·10

VOL. 6 956

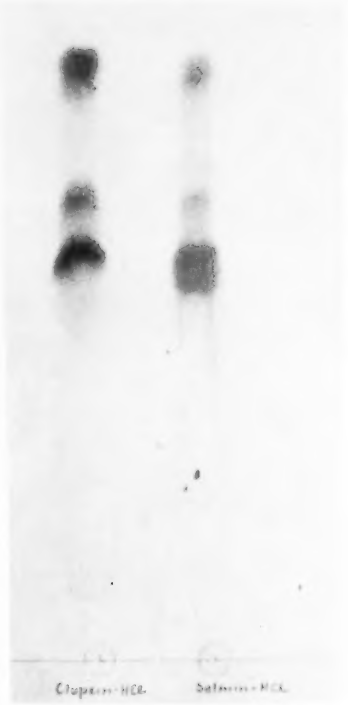


Fig. 3

Fig. 4

Fig. 5

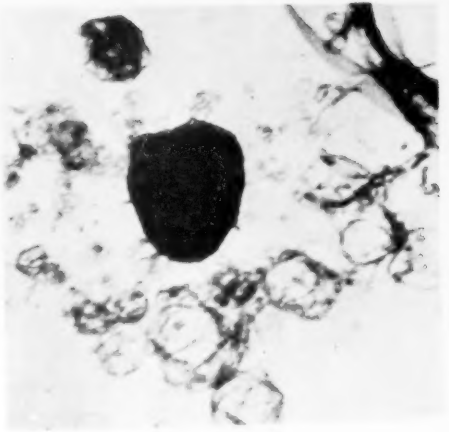


Fig. 6

Fig. 3. Separation of clupeine and salmine into several components by paper chromatography (see text).

Fig. 4. Trout sperm in Ringer solution. Osmium fixation.

Fig. 5. Trout sperm, suspended in distilled water for 10 min. Osmium fixation.

Fig. 6. Trout sperm, suspended in distilled water for 30 min. Osmium fixation.

Fig. 7. Trout sperm; isolated nuclei after repeated washings with distilled water. Osmium fixation.

Fig. 8. Trout sperm; Acetone dried nuclei. Osmium fixation.

VOL
6
195

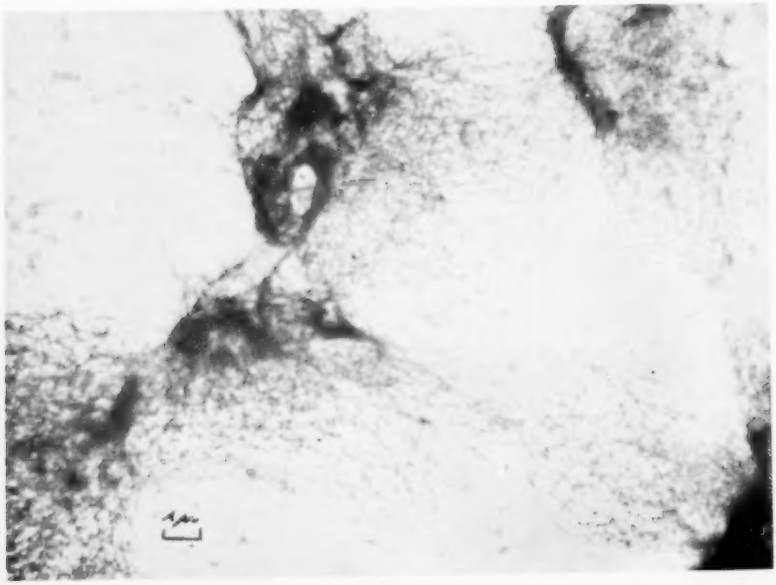


Fig. 9. Trout sperm: Nuclei treated on the blend with 10 per cent NaCl and subsequently washed with distilled water

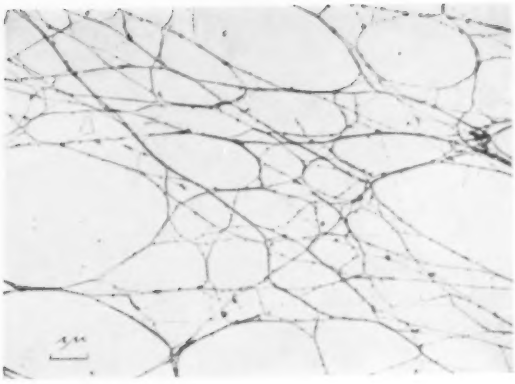


Fig. 10. Nucleoprotamine fibres: Trout sperm nuclei dissolved in 10 per cent NaCl and reprecipitated on the blend by distilled water.

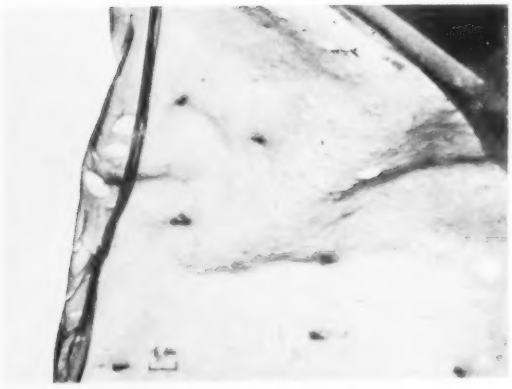


Fig. 11. Nucleoprotamine film. Prepared by spreading nucleoprotamine solution on water.

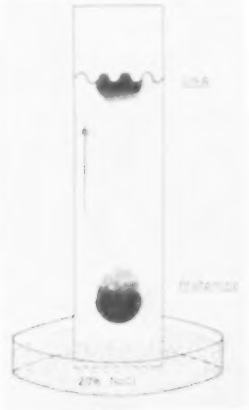


Fig. 12. Separation of nucleoprotamine into protamine and DNA by salting out. Chromatography on filter paper. (See text.)

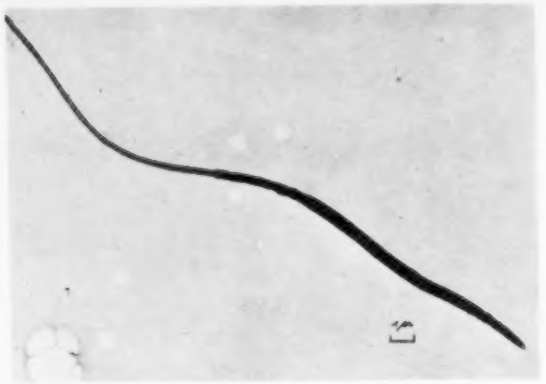


Fig. 13. Fowl sperm in Ringer solution. Osmium fixation.

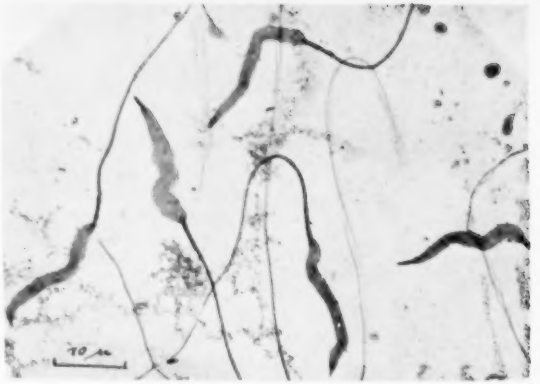


Fig. 14. Fowl sperm: Swelling of nuclei by gradual evaporation of Ringer solution on the blend. No preformed structures inside the nucleus. Insertion points of acrosome and tail become visible.

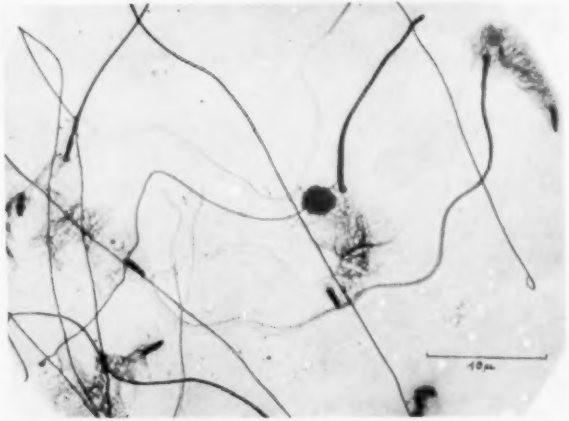


Fig. 15. Fowl sperm: Swelling of nuclei by gradual evaporation of Ringer solution on the blend. Nucleoprotamine reprecipitated by washing with distilled water.

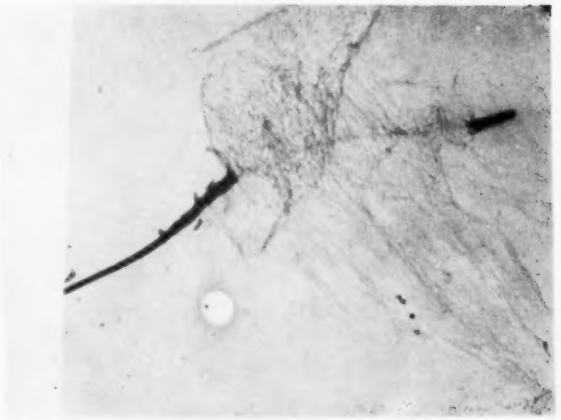


Fig. 16. Progressed state of nuclear dissolution.

VOL
6
195

Nucleoprotamines

Miescher already knew that the sperm nuclei of fish chiefly consist of nucleoprotamine. Mathew[60] and especially Steudel[61], [62], [63] investigated it more closely. The latter analyzed genuine nucleoprotamine and compared it with a synthetic substance prepared from protamine sulphate and thymonucleic acid. Both contained almost the same amount of nitrogen and phosphorus.[64]

Recently, it has become possible to extract nucleoprotamine cautiously with 10 per cent sodium chloride solution and precipitate it by diluting (as mentioned above). This method allows us to compare pure nucleoprotamine with the cell nuclei from which it was obtained, provided these were free of all cytoplasmic components.

For this reason we have taken special care in purifying the cell nuclei. We isolated them from fresh trout milt by plasmolysis in distilled water. Under the microscope one can observe how the cytoplasm surrounding the spermhead and tail swells, the fibrills form loops and finally both are detached from the nucleus. Only a slight mechanical impact (homogenizer) is needed for complete separation. The last traces of cytoplasm are removed by washing with ice cold water and centrifuging repeatedly. The naked nuclei are obtained as a pure white, loose sediment.

Figs. 4, 5, 6, 7, and 8 show the various stages of plasmolysis.

We wish to thank Dr. W. Lippert and Dr. O. Hug of the Max-Planck-Institute for Biophysics, Frankfurt a. Main, for having taken the electro-microscopic photographs.

According to the photographs, the nucleus does not change during this procedure either in volume, shape or optical density. In the electron microscope the nucleus appears to be a homogeneous compact and not transparent mass. By treating the nuclei with aceton and drying them over phosphorus pentoxide, the volume is hardly diminished; this means that they contain very little water.[65]

With this method we have prepared the nuclei of various trout species and of herring.[12], [22], [66] They only contain nucleic acid of the deoxytype: We never succeeded in finding ribonucleic acid. The reactions for tryptophane or sulphur-containing amino acids were always negative. When heated with Millon's reagent, their colour changed into light pink which disappeared before the boiling point was reached.

In 10 per cent chloride solution the nuclei dissolve completely. The solution is very viscous and only slightly opalescent. By pouring this solution into distilled water, the entire nuclear material is precipitated as a fine fibrous mass. The mother-lye is almost free of nitrogen and phosphorus.

Under suitable experimental conditions, the nucleoprotamines may

also be precipitated as a film on the optical blend of an electron microscope. Such preparations show minute particles within the nucleoprotamine film which remained undissolved. It is very probable that these represent parts of the proximal centriol located at the insertion-point of the spermatozoon tail in a heart-shaped inflection of the sperm nuclei. This is the only cytoplasmic constituent not always removed from the nuclei by plasmolysis. However, this only amounts to 1/1000 of the volume of a nucleus[65] (Figs. 9, 10, and 11).

We determined nitrogen, phosphorus and arginine in the nuclei as well as in repeatedly precipitated nucleoprotamine and, within the experimental error, found the same values for both materials. (See Table 8.)

TABLE 8

Analyses of Sperm Nuclei

	N	*P*	*N/P*	*Arginine*	*P/Arginine*
Salmo irideus . .	19·67	5·87	3·35	30·86	1:0·94
Salmo trutta . .	19·80	5·72	3·46	30·42	1:0·95
Salmo fontinalis .	19·78	5·88	3·36	30·44	1:0·92
Clupea harengus .	19·73	5·76	3·42	30·42	1:0·94

Analyses of Fibres of Nuclei

Samo irideus . .	19·52	5·65	3·43	30·60	1:0·96
Salmo trutta . .	19·67	5·71	3·44	30·20	1:0·94
Salmo fontinalis .	19·67	5·73	3·43	30·15	1:0·94
Clupea harengus .	19·57	5·68	3·44	30·77	1:0·96

The agreement in the values indicates that the entire nuclear material is contained in the precipitated nucleoprotamine. However, this does not mean that it still has the same structure, since the protamines and nucleic acids are separated from each other in sodium chloride solution. It can, moreover, hardly be expected that protamines and nucleic acid unite in the original manner, after having been precipitated in distilled water.

POLLISTER and MIRSKY[67] dialyzed such a solution of nucleoprotamines against 10 per cent sodium chloride and renewed the external solution repeatedly. Finally, nothing was left except nucleic acid; the protamines had dialyzed out. Protamines and nucleic acid can also be separated partially by centrifugation.

The following simple experiment demonstrates the dissociation even clearer: After a drop of nucleoprotamine solution is put on a stripe of filter paper and this is air-dried, a 20 per cent chloride solution is allowed

to ascend this stripe; the protamine remains at the starting-point while nucleic acid moves upward (Fig. 12).

The salt-like character of the nucleoprotamines is also indicated by chemical analysis. The last column of Table 8 shows that the ratio phosphorus : arginine is practically 1:1, or in other words: Every phosphoric acid residue is neutralized by a guanidine residue of the protamine component. Although this fact appears obvious, it is very difficult to construct a model of nucleoprotamine which is in accord with the present views on the structure of nucleic acids and proteins.[68], [59], [69], [70]

We also analyzed the nucleoprotamines of salmon and sturgeon. These investigations were performed on dried sperm heads instead of isolated nuclei, as mentioned above. Nevertheless, nucleosalmine showed the same relations between phosphorus and arginine. In nucleosturine, however, less than one molecule of arginine is available for one phosphoric acid residue because these are partially neutralized by the two other basic amino acids. In addition to the 0·76 equivalents of arginine, there are 0·19 equivalents of lysine and 0·15 of hisitidine.

TABLE 9

Analyses of Fibres of Nuclei

	N	*P*	*n/P*	*Arginine*	*P/Arginine*
Salmo salar	20·25	5·39	3·57	28·90	1:0·96
Acipenser sturio	19·72	5·47	3·60	23·51	1:0·76

According to our investigations, the nuclei of rooster spermatozoa also consist only of nucleoprotamine. Figs. 13–16 show that only the cell nucleus and nothing else is dissolved when spermatozoa are exposed to salt solutions of higher concentration.

Genetic Aspects

From these analyses it becomes evident that the genetic material of the sperm nuclei consists only of nucleoprotamine and the question is, how this can represent the divers characteristics which are transmitted from the fish father to his off-spring. Our analyses up to now suggest the following explanation: The protamines of every fish species differ in their composition or at least in the sequence of the amino acid in the peptide chain. Besides, every protamine consists of several components. Even if the nucleic acids in the sperm of different fish should prove to be the same, what we do not expect, the protamines guarantee sufficient variations. All according to the components which are combined with the DNA and their sequence along the nucleic acid, different nucleoprotamines result.

A rough calculation may demonstrate this. From the nitrogen and phosphorus content of the sperm nuclei and the nucleoprotamines it can be concluded that the ratio nucleic acid:protamine varies from 61/39 by herring to 67/33 by sturgeon. For the salmonidae it is 64/36.

According to preliminary determinations in the ultracentrifuge, the molecular weight of DNA is about 800,000; therefore, the molecular weight of nucleoprotamine is about $1{\cdot}2–1{\cdot}3 \times 10^6$.[29]

The molecular weight of the protamine components has not yet been determined exactly. It varies from 2000 to 12,000. If we assume an average of 5000, about 100 molecules of protamine are contained in 1 molecule of nucleoprotamine. If a protamine consists of ten different components, the number of combinations between protamine and DNA would be so large that each gene could, perhaps, be assured its own nucleoprotamine.

However, the question is, do the sperm nuclei contain enough nucleoprotamine molecules to realize these considerations. In answer to this question, we have determined the average nitrogen and phosphorus of a single cell nucleus and from this the DNA content was calculated. These determinations were carried out at different times and by different people and were confirmed repeatedly. The values for the different trout species and the herring vary between 6·4 and $7{\cdot}1 \times 10^{-9}$ mg (see Table 10).

TABLE 10

DNA content of a Single Sperm Nucleus of Various Fish Species expressed as mg $\times 10^{-9}$ *per Nucleus*

Calculated from	*P*	*N*
Salmo irideus .	6·4	6·7
Salmo fontinalis .	6·7	6·9
Salmo trutta . .	6·6	7·0
Clupea harengus .	6·9	7·1

If we assume the molecular weight of DNA to be 800,000, a single cell nucleus, as calculated by the Avogadro number, contains 5×10^6 molecules of DNA and, of course, the same number of nucleoprotamine molecules.

In connection with the figure given in Table 10, we wish to make the following remarks: According to our analyses the nuclei of fish sperm contain a greater amount of DNA than other authors have found in spermatozoa. VENDRELY,[71] e.g. found $2{\cdot}8–3{\cdot}4 \times 10^{-9}$ mg DNA in bull spermatozoa and MIRSKY and RIS[72] $1{\cdot}26 \times 10^{-9}$ mg DNA in rooster spermatozoa. However, these authors have applied methods for the isolation and determination by which losses can hardly be avoided.

It was, moreover, possible to confirm the above mentioned values by a physical method. We have determined the specific partial volume of the nucleoprotamine of herring and found that one gram in distilled water has a volume of 0·6947 cm^3. Provided that the molecular weight of nucleoprotamine is $1·3 \times 10^6$, a single molecule has a volume of $1·5 \times 10^{-18}$ cm^3 or $1·5 \times 10^{-3}$ $m\mu^3$. One sperm nucleus has a volume of $7\mu^3$ and room for $4·7 \times 10^6$ molecules of nucleoprotamine. That is the same amount of molecules we derived from the chemical data.

Since the electro-microscopic photographs show no structure within the nucleus, these 5 million nucleoprotamine molecules must lie very close together in a regular and definite order.

WILKINS and RANDALL[73] were so kind as to analyze by X-rays moist nuclei which we prepared from rainbow trout sperm. The X-ray diagrams indicate that the bundles of nucleoprotamine fibres in the nucleus have a three-dimensional, that is to say, a crystalline structure.

This structure is presumably characteristic for the different fish species and may give additional opportunities for variations besides those already provided by the protamines and perhaps the nucleic acids.

In the sperm nuclei of fish no reactions take place because they contain no enzymes and consume no oxygen.

This fact and the former considerations lead to the conclusion that the hereditary characteristics are essentially based on the structure of the sperm nucleus and that this small unit is subject to the same creative power which governs the development and formation of an organism.

Let us recall once more the simple means by which nature attains its great variety. The fish sperm nuclei contain only a few building blocks: seven amino acids on the one hand and 4–5 pyrimidine bases, deoxyribose and phosphoric acid on the other. The former yield protamines, the latter the different nucleic acids. When both unite, different nucleoprotamines result and these are arranged in a definite manner in space.

This was supported by grants from the Deutsche Forschungsgemeinschaft.

REFERENCES

(1) MIESCHER, F.; *Die histochemischen und physiologischen Arbeiten, Leipzig* 1897
(2) KOSSEL, A.; *Protamine und Histone*, Leipzig 1929
(3) DALY, M. M., MIRSKY, A. E. and RIS, H.; *J. Gen. Physiol.* 34 (1951) 439
(4) FISCHER, H. und KREUZER, L.; *Z. Physiol. Chem.* 293 (1953) 176
LIPPERT, W., HUG, O., WAGNER, L. und FISCHER, H.; Chromosoma, in press
(5) VENDRELY, R. and VENDRELY, P.; *Nature* 172 (1953) 30
(6) DALLAM, R. D. and THOMAS, L. E.; *Biochem. et Biophys. Acta* 11 (1953) 79
(7) WALDSCHMIDT-LEITZ, E., ZIEGLER, E. F., SCHAEFFNER, A. and WEIL, L.; *Z. physiol. Chem.* 197 (1931) 219
(8) FELIX, K. and DIRR, K.; *Z. physiol. Chem.* 184 (1929) 111
(9) FELIX, K. and MAGER, A.; *Z. physiol. Chem.* 249 (1937) 111

VOL. 6 1956

(10) RAUEN, H. M., STAMM, W. and FELIX, K.; *Zs. physiol. Chem.* 291 (1952) 275
(11) BLOCK, R. J., BOLLING, D., GERSHON, H. and SOBER, H. A.; *Proc. Soc. exp. Biol. a. Med.* 70 (1949) 494
(12) FELIX, K., FISCHER, H., KREKELS, A. and MOHR, R.; *Z. pyhsiol. Chem.* 287 (1951) 224
(13) MIRSKY, A. E. and POLLISTER, A. W.; *Proc. Nat. Acad. Sci. U.S.A.* 28 (1942) 344
(14) FELIX, K., FISCHER, H., KREKELS, A. and RAUEN, H. M.; *Z. physiol. Chem.* 286 (1950) 67
(15) KURODA, Y.; *J. Biochem.* (Japan) 38 (1951) 115
(16) FELIX, K., FISCHER, H. and KREKELS, A.; *Z. physiol. Chem.* 289 (1952) 127
(17) TRISTAM, G. R.; *Nature* 160 (1947) 637
(18) HAMER, D. H. and WOODHOUSE, D. L.; *Nature* 163 (1949) 689
(19) VELICK, S. F. and UDENFRIEND, S.; *J. Biol. Chem.* 191 (1951) 233
(20) CORFIELD, M. C. and ROBSON, A.; *Biochem. J.* (Brit.) 55 (1953) 517
(21) MILLS, G. L.; *Biochem. J.* (Brit.) 50 (1952) 707
(22) FELIX, K., FISCHER, H., KREKELS, A. and MOHR, R.; *Z. physiol. Chem.* 289 (1951) 10
(23) ANDO, T., ISHII, S., HASHIMOTO, C., YAMASAKI, M. and IWAI, K.; *Bull. chem. Soc. Japan* 25 (1952) 132
(24) ANDO, T., IWAI, K., YAMASAKI, M., HASHIMOTO, C., KIMURA, M., ISHII, S. and TAMURA, T.; *Bull. chem. Soc. Japan* 26 (1953) 406
(25) HULTIN, T. and HERNE, R.; *Arkiv. Kemi. Mineral. Geol.* 26 A, Nr. 20 (1949)
(26) RASMUSSEN, E. and LINDERSTRÖM-LANG, K.; *Z. physiol. Chem.* 22 (1934) 97
(27) RAUEN, H. M., STAMM, W. and FELIX, K.; *Z. physiol. Chem.* 292 (1953) 101
(28) FELIX, K. and STAMM, W., unpublished
(29) DAIMLER, B.; *Kolloidz.* 127 (1951) 8, 97
(30) ZAHN, R. K. and STAMM, W.; *Naturwiss.* 41 (1954) 95
(31) FELIX, K. and KREKELS, A., to be published
(32) WALDSCHMIDT-LEITZ, E. and VOH, R.; *Z. physiol. Chem.* 298 (1954) 257
(33) SANGER, F.; *Biochem. J.* 39 (1945) 507
(34) WALDSCHMIDT-LEITZ, E., KÜHN, K. and ZINNERT, F.; *Experientia* 7 (1951) 183
(35) FELIX, K., RAUEN, H. M. and ZIMMER, G. H.; *Z. physiol. Chem.* 292 (1952) 228
(36) FELIX, K. and KREKELS, A.; *Z. physiol. Chem.* 295 (1953) 107
(37) PORTER, R. R. and SANGER, F.; *Biochem. J.* 42 (1948) 287
(38) EDMAN, P.; *Acta chem. Scand.* 4 (1950) 283
(39) HÜBNER, L.; Dissertation, Frankfurt a. Main (1954)
(40) ABDERHALDEN, E. and BROCKMANN, H.; *Biochem. Z.* 225 (1930) 386
(41) WESSELY, F.; *Nature* 169 (1952) 708
(42) FELIX, K.; *Ber. Physiol.* 61 (1931) 349
(43) DIRR, K. and FELIX, K.; *Z. physiol. Chem.* 205 (1932) 83
(44) BERGMANN, M. and KÖSTER, H.; *Z. physiol. Chem.* 159 (1926) 179
(45) AKABORI, Sh., OHNO, K. and NAVITA, K.; *Bull. Chem. Soc. Japan* 25
(46) WALDSCHMIDT-LEITZ, E. and GAUB, K.; *Chem. Ber.* 85 (1952) 352
(47) FELIX, K., INOUYE, K. and DIRR, K.; *Z. physiol. Chem.* 211 (1932) 187; *Z. physiol. Chem.* 293 (1953) 10
(48) DIRR, K. and FELIX, K.; *Z. physiol. Chem.* 209 (1932) 5
(49) FELIX, K., HIROHATA, R. and DIRR, K.; *Z. physiol. Chem.* 218 (1933) 269
(50) FELIX, K., RAUEN, H. M. and ZIMMER, G. H.; *Z. physiol. Chem.* 291 (1952) 228
(51) FELIX, K. and MEISSNER, P.; Dissertation, Frankfurt (1953)

VOL
6
195

REFERENCES

(52) MARSHAK, A. and VOGEL, H. J.; *J. Biol. Chem.* 189 (1951) 597
(53) WYATT, G. R.; *Biochem. J.* 48 (1950) 584
(54) FELIX, K. and JILKE, J.; unpublished results, Dissertation, Frankfurt a. Main (1953)
(55) ZAMENHOFF, S., SHETTLES, L. B. and CHARGAFF, E.; *Nature* 165 (1950) 756
(56) DALY, M. M., ALLFREY, V. G. and MIRSKY, A. E.; *J. Gen. Physiol.* 33 (1950) 497
(57) WYATT, G. R.; *Biochem. J.* 48 (1951) 581
(58) CHARGAFF, E., VISCHER, E., DONIGER, R., GREEN, G. and MISANI, F.; *J. Biol. Chem.* 177 (1949) 405
(59) WATSON, J. D. and CRICK, F. H. C.; *Nature* 171 (1953) 737
(60) MATHEW, A.; *Z. physiol. Chem.* 23 (1897) 399
(61) STEUDEL, H.; *Z. physiol. Chem.* 72 (1911) 305
(62) STEUDEL, H.; *Z. physiol. Chem.* 73 (1911) 471
(63) STEUDEL, H.; *Z. physiol. Chem.* 83 (1912) 72
(64) STEUDEL, H. and PEISER, E.; *Z. physiol. Chem.* 122 (1922) 298
(65) FISCHER, H., HUG, O. and LIPPERT, W.; *Chromosoma* 5 (1952) 69
(66) FELIX, K. and KREKELS, A.; *Z. physiol. Chem.* 293 (1953) 284
(67) POLLISTER, A. W. and MIRSKY, A. E.; *J. Gen. Physiol.* 30 (1947) 101
(68) PAULING, L. and COREY, R. B.; *Nature* 171 (1953) 346
(69) WILKINS, M. H. F., STOKES, A. R. and WILSON, H. R.; *Nature* 171 (1953) 740
(70) FRANKLIN, R. E. and GOSLING, R. G.; *Nature* 171 (1953) 745
(71) VENDRELY, C.; *Bull. Biol.* 86 (1952) 1
(72) MIRSKY, A. E. and RIS, H.; *Nature* 163 (1949) 666
(73) WILKINS, M. H. F. and RANDALL, J. T.; *Biochimica et Biophysica Acta* 10 (1953) 192

VOL
6
195

2

THE STRUCTURE OF CHROMOSOMES

E. J. Ambrose

CONTENTS

	Page
I. INTRODUCTION	26
II. THE CHEMICAL COMPOSITION OF NUCLEAR POLYMERS	26
(*a*) Proteins	26
(*b*) Nucleic acids	28
(i) Deoxyribose nucleic acid	28
(ii) Ribose nucleic acid	28
III. THE MOLECULAR STRUCTURE OF NUCLEAR POLYMERS	29
(*a*) Nucleic acid	30
(*b*) Nucleo-protein	30
IV. THE COLLOIDAL PROPERTIES OF NUCLEO-PROTEIN	32
Effects of concentration	33
V. THE STRUCTURE OF CHROMOSOMES	34
(*a*) Sperm heads	34
(*b*) Salivary gland chromosomes	34
(*c*) The resting nucleus of normal cells	40
VI. CHROMOSOMES IN DEVELOPMENT	41
VII. CHROMOSOMES IN DIVIDING CELLS	43
(*a*) The synthesis of chromosomes	43
(*b*) Chromosomes in division	48
REFERENCES	52

2

THE STRUCTURE OF CHROMOSOMES

E. J. Ambrose

I. Introduction

The structure and behaviour of chromosomes in dividing cells has been the subject of study by cytologists for nearly 100 years. It would not be possible to cover this vast subject in a brief review. But in recent years much work has been carried out on the physical chemistry of nucleic acids and nucleo-proteins; new methods have also become available for the study of structure within living cells.

In this article an attempt has been made to correlate some of the observations obtained by the two methods, particularly in relation to an interpretation of chromosome behaviour through the successive stages of the cell cycle.

II. The Chemical Composition of Nuclear Polymers

The chemical composition and structure of the proteins and nucleic acid of the cell nucleus have been described in Vol. IV of this series by Davison, Conway and Butler.[1] For the purposes of the present article it will only be necessary to recapitulate a few of the structural aspects which are closely related to the studies of actual chromosomes and to provide some notes on more recent work.

(a) *Proteins*

The composition and structure of the protamines are considered in this volume by Felix, Fischer and Krekels.[2] These highly basic proteins of low molecular weight (4000–10,000) are replaced in tissues other than sperm heads by the histones, which are also of comparatively low molecular weight (14,000–20,000). Recent developments in the study of the proteins of the nucleus have been mainly concerned with attempts to fractionate and to discover some measure of cell specificity, i.e. differences in the histones isolated from various tissues. For example, Davison has carried out fractionations and has found that histones extracted by strong acids contain originally two electrophoretically distinct components. However, in neutral solutions, marked complex formation between the two takes place, leading to the appearance of a third peak intermediate between the other two. (In some cases aggregation, leading to insoluble products occurs.) Davison, Butler, James and Shooter[3] were also able to obtain two electrophoretically homogeneous substances by the following fractionation. After partial

precipitation, with ammonia, the residue was found to be electrophoretically equivalent to the fast component; while extraction of the mixed nucleo-protein with a weak acid (citric acid) gave the other (slower) component. Amino acid analyses of these substances showed great differences; the former being rich in lysine and the latter in arginine.

TABLE I

	Slow component		*Fast component*
	Fractionated by ammonia-acetone	*Extracted by citric acid*	*Fractionated by sodium chloride*
Alanine	21·0	21·6	8·0
Arginine	4·1	3·6	15·6
Aspartic acid	3·0	4·0	5·1
Glutamic acid	2·9	5·0	9·7
Glycine	5·1	4·7	5·1
Histidine	0	0	3·9
Isoleucine	1·4	1·8	6·6
Leucine	3·6	5·0	9·4
Lysine	33·0	23·1	9·5
Methionine	0	0	1·9
Phenylalanine	0	0	3·7
Proline	8·9	7·2	2·1
Serine	5·4	4·1	2·2
Threonine	6·0	5·8	5·9
Tyrosine	1·0	1·0	5·0
Valine	5·0	4·8	6·3

A lysine rich fraction has also been obtained by DALY and MIRSKY.[4] CRAMPTON and MOORE[5] have effected a fractionation of complete histone on a resin column. The histones are not normally eluted from the columns, but CRAMPTON and MOORE obtained the two principal fractions and several subsidiary peaks by elutriation with 2 M barium acetate.

In so far as the detection of tissue or cell specificity is concerned, some evidence on this point has been obtained by STEDMAN and STEDMAN[6] who found differences in the total arginine content of basic proteins in fowl erythrocytes as compared with thymocytes.

But in the case of mammalian cells, CROFT, MAURITZEN and STEDMAN were only able to demonstrate physical differences. They found that the extrapolated isoelectric points of beef thymus and liver occurred at *p*H 11·0 and 10·8, respectively. They therefore proposed another test for identity; when dialysed together they gave one peak in the electrophoresis apparatus, but when dialysed separately and mixed just before electrophoresis two peaks were obtained.

It is known that histones undergo changes on prolonged dialysis, possibly due to proteolytic enzymes, and the work of DAVISON *et al.* indicates that complex formation readily occurs between histones from the same tissue. It is likely that much work needs to be done before the cell specificity of histones can be systematically investigated.

In addition to the basic proteins of the nucleus a small amount (never more than 5 per cent) of a non-basic protein has been detected in cells other than sperm heads by STEDMAN and STEDMAN.[8] Little new work on this material has been carried out during the past two years.

(b) *Nucleic acids*

(i) *Deoxyribose nucleic acid*—Much work on the chemical constitution of nucleic acid has been stimulated by the CRICK and WATSON[9] double spiral model already described in this series (DAVISON, CONWAY and BUTLER[1]), which requires that the bases adenine-guanine and thymine-cytosine should occur in such proportions that the ratios A/G and T/C should be equal to 1. Values very near to unity were obtained by CHARGAFF, VISCHER and ZAMENHOF.[10] The more recent analyses appear to be in general agreement with the earlier conclusions.

As in the case of histones, efforts are now being concentrated on attempts to demonstrate specificity in nucleic acids by fractionation procedures. CHARGAFF, CRAMPTON and LIPSHITZ[11] obtained a partial fractionation by successive extractions with salt solutions of increasing strength. They thought that this was the result of breaking salt like bonds between the DNA and the histone particles. However, LUCY and BUTLER[12] found that the same result could be achieved by successive extractions with the same salt solution. These workers have also found that the first fractions of DNA which are richer in *guanine* are also accompanied by a histone which is richer in *lysine*; so that it would appear likely that the different DNA fractions are combined specifically with different histone fractions. These results are clearly of considerable relevance to the problem of genetic character as expressed by the nucleo-protein structures of the chromosome.

(ii) *Ribose nucleic acid*—Ribose nucleic acid is contained mainly in the nuclear inclusion known as the nucleolus. In these acids the thymine (5-methyl uracil) is replaced by uracil, while the sugar is ribose instead of deoxyribose. Recent work has shown that the pairs of bases adenine/uracil and guanine/cytosine do not show such a constant one to one relationship as in deoxyribose nucleic acid. In the nucleoli of starfish oocytes VINCENT[13] finds $A/U = 1{\cdot}47$ and $G/C = 1{\cdot}34$. ELSON and CHARGAFF[14] find in the whole nuclei of rat liver, ox liver and ox kidney $A/U = 0{\cdot}85 - 1$. $G/C = 0{\cdot}88$.

The molecular weight of ribose nucleic acid is considerably lower than that of deoxyribose nucleic acid. POLLISTER and LEUCHTENBERGER[15] have in fact been able to show that 25 per cent of the nuclear basophilia (test for RNA) can be removed by a cold solution of 5 per cent trichloracetic acid, in which ribose nucleic acid of the accepted molecular weight is not soluble. Low molecular weight material is evidently present in the nucleolus.

III. The Molecular Structure of Nuclear Polymers

There are certain properties of proteins and nucleic acids which are particularly relevant to chromosome behaviour and structure. These involve the arrangement of the long chain polymers and characteristic folding processes, which can take place within the molecule.

Since the discovery of the folded (α) and extended (β) forms of chain in proteins (ASTBURY and STREET,[16] ASTBURY and WOODS[17]), a synthetic approach has shown that polypeptide chains possess two characteristically stable forms. In the first form (β) the chains are extended and N—H and C=O bonds lie approximately at *right* angles to the chain axis (Fig. 1*a*). N—H . . . O=C hydrogen bonds are formed *between* the chains (*inter*-chain). These bonds have about $\frac{1}{10}$ of the strength of covalent bonds and can break and reform reversibly. The second form of chain is shown in Fig. 1*b*. In this case the hydrogen bonds are formed internally (*intra*-chain) by ring formation, a structure which was discovered experimentally by the use of polarized infra-red radiation. (AMBROSE and HANBY;[18] AMBROSE and ELLIOTT.[19]) X-ray studies were made by BAMFORD, HANBY and HAPPEY.[20] More recent work, using mainly X-ray diffraction methods, has suggested that the rings may be larger than those shown in Fig. 1*b*, (13 membered instead of seven membered) (PAULING, COREY and BRANSON;[21] PAULING and COREY;[22] PERUTZ;[23] BAMFORD, BROWN, ELLIOTT, HANBY and TROTTER[24]). This leads to a helical form of chain (α-helix) as shown schematically in Fig. 1*c*.

Which of the two forms of chain occurs in basic proteins? Some spectroscopic evidence is available on this point. A synthetic polypeptide polylysine (WALEY and WATSON[25]), has been examined in polarized infra-red radiation (Ambrose unpublished) and is found to be stable in the folded form of Fig. 1 (*b* or *c*). Unfortunately the protamines have molecular weights which are too low for the preparation of orientated fibres, but an examination of the infra-red frequencies of the peptide absorption bands indicate the presence of a folded chain in this material also. (1654–61 cm^{-1}, 1531–51 cm^{-1} for the folded and 1628–32 cm^{-1}, 1521–26 cm^{-1} for the β form, AMBROSE and ELLIOTT;[19] ELLIOTT.[26]) It may be concluded that, in the free state, basic proteins tend to assume the folded configuration of the polypeptide chain.

(a) *Nucleic acid*

The CRICK and WATSON[9] model for nucleic acid consisting of twin spiral phosphate ester chains with paired bases attached by hydrogen bonds in the centre of the helix is shown in Fig. 3, Vol. 4, p. 169, of this series (DAVISON, CONWAY and BUTLER). This model seems at present to provide a good working hypothesis for the structure of deoxyribose nucleic acid. The hydrogen bonds are formed between the paired

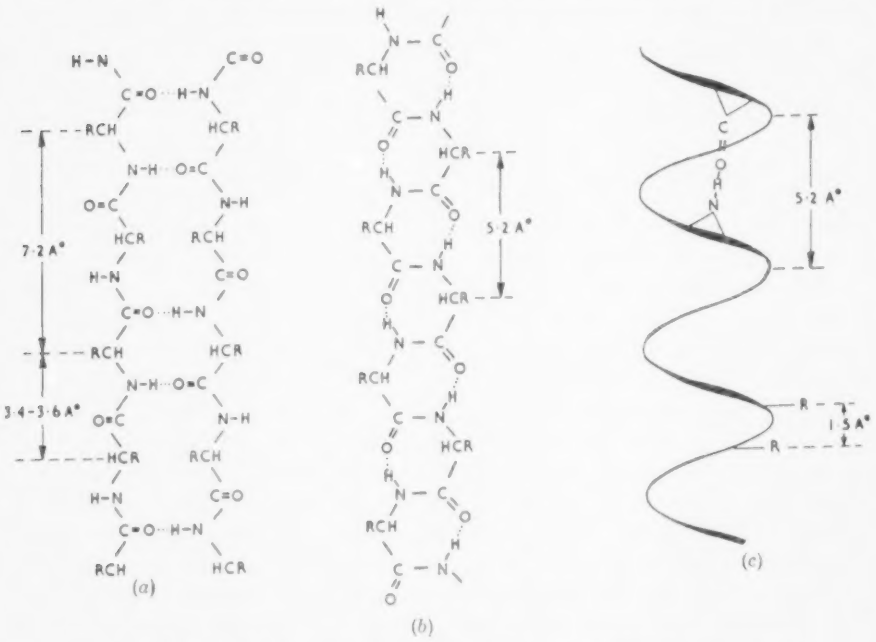


Fig. 1. (*a*) Extended form of polypeptide chain; . . . hydrogen bonds lying at right angles to chain axis. (*b*) Folded form of polypeptide chain with seven membered rings (AMBROSE and HANBY); . . . hydrogen bonds lying parallel to chain axis. (*c*) Folded form of polypeptide chain (PAULING, COREY and BRANSON) with 13 membered rings; . . . hydrogen bonds lying parallel to chain axis.

molecules so that there is, from the physico-chemical point of view, a resemblance to the folded form of polypeptide chain in which the hydrogen bonds are *intra*-molecular. The structure of nucleic acid is known experimentally to be moderately rigid (SADRON,[27] POUYET,[28] CONWAY and BUTLER[29]), in agreement with a model of this type.

(b) *Nucleo-protein*

In nucleo-protein extracted from cell nuclei, the forces of cohesion between protein and nucleic acid can be broken by salt solutions above a

concentration of 0·5 M. The basic problem arises as to whether the molecular configuration of nucleic acid is affected by salt-like combination with basic protein or vice versa. WILKINS and RANDALL[30] studied the X-ray diffraction patterns of fresh sperm heads and found that the reflections corresponded to spacings in the molecule which were similar to those observed in the paracrystalline form of nucleic acid alone (the form which is stable in the presence of more than 30 per cent of water, RILEY and OSTER[31]). The small scale structure of nucleic acid does not appear therefore to be affected by combination with protein. Infra-red studies of nucleic acid (FRASER and FRASER[32]) and of nucleoprotein (AMBROSE and BUTLER[33]) support this view. In both cases the C=O bonds of the purine and pyrimidine bases lie at right angles to the fibre axis. WILKINS[34] has examined fibres of nucleic acid in polarized ultra-violet radiation. He finds that, in the paracrystalline form, the fibres show strong absorption when the *E*-vector of the polarized beam lies *perpendicular* to the fibre axis, as is to be expected if the rings of the purine and pyrimidine bases lie at right angles to, or approximately at right angles to the chain axis. Here again WILKINS has observed similar ultra-violet dichroism with sperm heads.

Certain dyes such as toluidine blue and neutral red are adsorbed on nucleic acid (WHITE and ELMES[35]). These fibres when observed in polarized light, show strong absorption when the *E*-vector lies perpendicular to the fibre axis, but weak absorption when it is parallel to the axis. This appears to be due to an orderly deposition of the dye molecules on the parallel chains of nucleic acid. Here again a similar type of dichroism can be observed with nucleoprotein fibres (AMBROSE and GOPAL-AYENGAR[36]). The optical properties of nucleo-protein are summarized in Fig. 2*b*, *c* and *d*.

Taking all the results described above into account, it is clear that there is a considerable weight of evidence suggesting that the small scale structure of nucleic acid is not affected by combination with protein.

ASTBURY pointed out that the periodicity of 3·4 Å along the fibre axis in nucleic acid could be due to a packing of the flat rings of the bases on top of each other like a pile of pennies. Subsequent models for nucleic acid including that of CRICK and WATSON have taken over this concept. ASTBURY[37] pointed out that this periodicity was similar to that of β proteins, and that a β protein (Fig. 1*a*) might be associated with nucleic acid. The only fresh evidence on this point has been obtained by AMBROSE and BUTLER[33] and by AMBROSE (unpublished).

When clupein is combined with nucleic acid at least a portion of the peptide absorption band is shifted to a frequency corresponding to the β form. In orientated films of herring sperm nucleoprotein, this frequency is also observed and the dichroism indicates that the peptide N—H . . . O=C hydrogen bonds lie approximately at *right angles*

to the fibre axis. In the case of tobacco mozaic virus, FRASER[38] has found a frequency corresponding to a folded chain; but the amount of nucleic acid in this virus is very small (5 per cent) and it would not be expected that the bulk of the protein would have its configuration affected by combination with nucleic acid, as in the case of nuclear materials, which contain about equal proportions of nucleic acid and protein. The infra-red evidence therefore suggests that basic proteins,

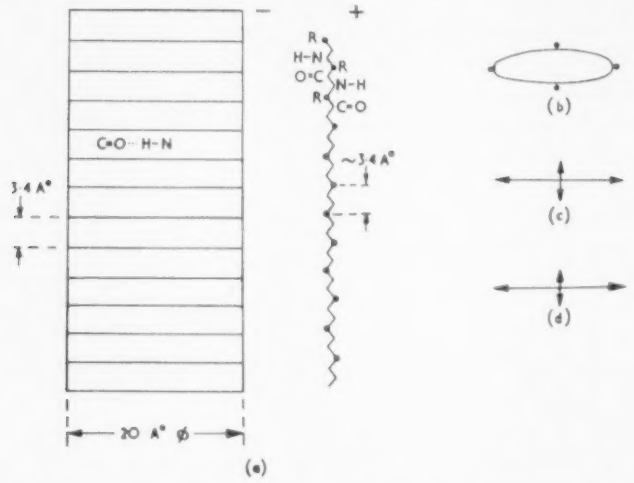


Fig. 2. (*a*) Diagrammatic representation of nucleo-protein molecule with purine and pyrimidine rings packed on top of each other, with their $C=O$ bonds $\perp r$ to chain axis; also extended polypeptide chain associated with nucleic acid. (*b*) Shape of polarizability ellipsoid (double refraction) for nucleic acid and nucleo-protein. (*c*) Vectors showing intensity of ultra-violet absorption in polarized light. (*d*) Vectors showing intensity of visible light absorption in polarized light, after staining with neutral red.

which are stable in the folded form, may be transformed into the β form when in contact with nucleic acid (Fig. 2*a*).

IV. The Colloidal Properties of Nucleo-protein

The colloidal properties of nucleo-protein have been described in detail in the previous article. For example, AMBROSE and BUTLER showed that films of nucleo-protein show a minimum swelling at a physiological salt concentration (0·14 M NaCl). The minimum swelling corresponds to a minimum electrophoretic velocity of the gel particles, which are negatively charged. The electrical properties of the nucleo-protein are in fact largely dependent upon this negative charge, because the swelling is not affected by the nature of the anion. It is highly dependent upon the valency of the cation. Concentrations of 0·001 M Ca or Mg ions will almost completely inhibit swelling.

Nucleo-proteins are also dependent upon changes of hydrogen ion concentration. Strong alkali causes dissociation of the nucleic acid and protein. Acid treatment can lead eventually to denaturation, and complete insolubility. But over a smaller range of *p*H change, an interesting phenomenon is observed in the case of the T_2 phage virus. This virus is of interest in that it is composed of deoxyribose nucleic acid and protein in almost equal proportions by weight, i.e. it has a general chemical composition rather similar to that of the herring sperm and thymus nucleo-proteins. This does not imply that the detailed structure is in any sense the same, but we might expect to find certain resemblances in the general colloidal properties. AMBROSE and LOVELESS (AMBROSE[39]) have found that on reducing the *p*H of a buffered solution of the virus from 6·8 to 4, extremely long fibres of uniform diameter ($\sim 1\,\mu$) are produced. These fibres show birefringence and can be stained with neutral red and toluidine blue, when they show dichroism similar to that described in the case of the stretched herring sperm nucleo-protein or thymus nucleo-protein. They are elastic and can be stretched reversibly to over twice their original length. The original solution of virus contains particles of diameter 800 Å and tail length 1000 Å.

Nucleic acid alone, when treated under the same conditions, did not give rise to fibres, but rather to an unorientated granular precipitate. The effects of salt and of hydrogen ion concentration upon nucleo-protein therefore indicate that changes in the electrolytic composition of the cell could have a marked effect upon the type of structure observable within the nucleus.

Effects of concentration

In the case of solutions of certain synthetic polypeptides, which are allowed to concentrate slowly under very still conditions, regions showing marked molecular orientation have been observed (ELLIOTT and AMBROSE[40]). The solution separates into a two phase system, which sometimes assumes the form of spherulites. The phase which has the higher concentration of polymer is *spontaneously* birefringent.

A somewhat similar experiment can be carried out with solutions of nucleo-protein and nucleic acid. It is necessary to use salt-free solutions in order to avoid changes of salt concentration. In a concentrating solution of calf thymus nucleo protein, prepared according to the method described by BUTLER and DAVISON,[41] the formation of orientated regions also takes place. In this case, birefringent fibres are observed, which can also be stained with neutral red and toluidine blue, when they show dichroism similar to that of the virus fibres. Physico-chemical evidence in favour of an association process at a critical concentration of 1–3 per cent has been obtained. Here again nucleic

acid alone does not show the same effects; if fibres are produced they are below the resolving power of the light microscope and there is a continuous transition to the birefringent regions at very high concentrations, with no sharp boundaries.

The effects of electrolytes and of changes of concentration upon the nuclear structure in living cells are considered in the following sections.

V. The Structure of Chromosomes

We will now consider the organization of the nuclear polymers, as it has been observed to occur within the intact cell.

(a) *Sperm heads*

Sperm heads presumably represent a late stage in the development of the resting nucleus and contain mature chromosome threads because their combination with the egg nucleus initiates division of the egg.

The optical properties of elongated sperm heads have been described by Schmidt.[42] Like nucleic acid fibres, they are negatively birefringent with reference to the long axis. In spherical heads there is no birefringence. As already mentioned, these results have been correlated with x-ray diffraction studies by Wilkins and Randall.[30] These investigators found that, with elongated sperm heads, a diffraction pattern showing orientation and spacings similar to those of nucleic acid was observed. The pattern in spherical heads also corresponded to similar spacings, although no preferred orientation could be seen. The pattern was more diffuse than that obtained with pure nucleic acid.

These results suggest that the small scale molecular structure in the sperm heads is similar to that obtained in the model systems already described, and that in the elongated sperm the molecules are constrained into a parallel orientation by the form of the enveloping membrane.

(b) *Salivary gland chromosomes*

During the growth of Dipteran larvae, cell division is arrested in certain tissues but the cells continue to grow, reaching in some cases a diameter of $\sim 300\ \mu$, in contrast to normal tissue cells which have a diameter of $\sim 15\ \mu$. The chemical composition and structure of the chromosomes has been the subject of many investigations since their discovery in 1881 by Balbianni.[43] Studies of the chemical composition of the chromosomes have been carried out almost entirely on fixed cells and for this reason must be treated with some reserve. The general appearance of a salivary gland chromosome is shown diagrammatically in Fig. 3. The long rope-like chromosomes are marked by a series of dense transverse bands. The bands are of varying shape and dimensions and are arranged in an aperiodic sequence. There are ~ 1000 bands in the x-chromosome of D. Melanogaster. Engström and Ruch[44] have

measured the mass of the band and interband regions, using absorption by soft x-rays. They find that the mass in the bands is considerably greater than in the inter-band regions. The chemical composition has been studied by ultra-violet absorption methods. (CASPERSSON[45]) The bands contain a high proportion of nucleic acid (10–30 per cent) and proteins with basic side chains (10–15 per cent). The interband regions do not contain nucleic acid; their proteins are not noticeably basic and contain 2–5 per cent of tyrosine and tryptophane.

The character and position of the bands has been correlated with the genetic character by numerous cytogenetic studies. But in addition

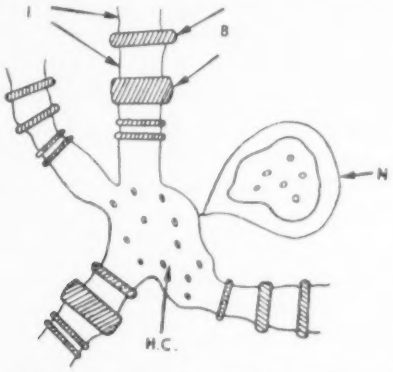


Fig. 3. Salivary gland chromosomes. *B*—Bands. High mass, high nucleic acid content. DNA with some RNA? Contain basic amino acids. *I*—Inter-band regions. Low mass. Little or no nucleic acid. No basic aminoacids. *HC*—heterochromatin. Less organized. Much nucleic acid (DNA with RNA?) Mainly basic protein only. *N*—Nucleolus. Ribosenucleic acid, possibly some acidic and some basic protein.

to the bands more diffuse and less organized regions are observed along the chromosome thread. These are the heterochromatic regions which have been studied particularly by SCHULTZ.[46] They do not show the characteristic mutation phenomena which are correlated with the dense transverse bands. But they do appear to play a part in deciding processes of growth and early development (GOLDSCHMIDT[47]). They appear to be related to the metabolism of nucleic acid and protein synthesis and can be developed to varying extents in different tissues (PATTERSON, LANG, DACKERMAN and SCHULTZ[48]). The heterochromatic regions contain nucleic acid and basic protein but they appear to contain only a small amount of the non-basic protein of the interband regions. The chromosome threads of the different chromosomes are distinct except at the heterochromatic regions where they often associate to form a *chromocentre*. The only other distinct organelle in the salivary gland nucleus is the nucleolus. This is a spherical or rather irregularly shaped body which appears to consist of two regions, an

outer and an inner region (Fig. 3). The inner region has a distinctly vacuolated appearance in the living state. This region of the cell contains ribose nucleic acid and according to CASPERSSON, a basic protein.

In addition to the ultra-violet studies, observations on the fixation of dye molecules and the effect of enzymes have been made (SCHULTZ[49]). The bands are Feulgen positive (i.e. they give a reaction for deoxyribose nucleic acid) and they stain with basic dyes. Both these reactions are destroyed by deoxyribonuclease. BRACHET[50] states that there are inclusions of ribose nucleic acid, as indicated by the methyl green-pyronin stain in Chironomus salivary gland chromosomes, in addition to the nucleoli. DANIELLI[51] found that the bands contained histidine, tyrosine and tryptophane, but that the inter-band regions were deficient in these materials although they contained amino-groups.

The basic *structural* element of the thread appears to be protein according to MAZIA and JAEGER.[52] They found that the chromosomes were digested by buffered trypsin (not by pepsin which does not attack histones). A similar conclusion was reached by YASUZUMI, ODATE and OTA in their electron microscope investigations. MCDONOUGH, ROWEN and MOHN found that the microscopic structure and general appearance of the salivary gland chromosomes of D. Melanogaster was not dependent upon the presence of deoxyribose nucleic acid. But KAUFMANN, GAY and MCDONALD have carried out a systematic investigation using digestion with pure enzymes and staining procedures. They find that part of the DNA is associated with histone while part of the RNA is associated with a tryptophane containing protein (probably acidic) and conclude that the whole chromosome must be considered as an integrated structure with no single structural element. KAUFMANN[56] has recently examined the effect of trypsin on freshly isolated salivary gland chromosomes and finds that unfixed chromosomes are little attacked by trypsin, although fixed chromosomes are readily attacked. (It is of course generally considered that denaturation of the protein is the first mode of attack by proteolytic enzymes.)

We will now consider some of the more basic aspects of the structure of salivary gland chromosomes. Certain of the optical properties of the chromosomes are illustrated in Fig. 4*e* and *f*. The chromosomes are weakly birefringent in the banded regions (SCHMIDT[42] and PFEIFFER[57]). The sign is negative, when referred to the length of the chromosome. It is not normally possible to detect dichroism on staining with low concentrations of vital stains (stains which are not immediately harmful to living cells) but AMBROSE and GOPAL-AYENGAR[36] using a sensitive instrument, based on the principle of the Flicker photometer (Flicker Dichroscope), were able to detect dichroism when a critical concentration of dye had been reached. This effect also showed the same

VOL
6
195

sign as that observed with orientated films of nucleo-protein. RUCH[58] has also detected ultra-violet dichroism in the salivary gland chromosomes. (The sign is not stated.)

The optical properties of salivary gland chromosomes therefore suggest that there is a small degree of molecular orientation of nucleic acid within the structure. Further information about the basic structural arrangement is given by swelling experiments. In the case of the orientated films of nucleo-protein already described (AMBROSE

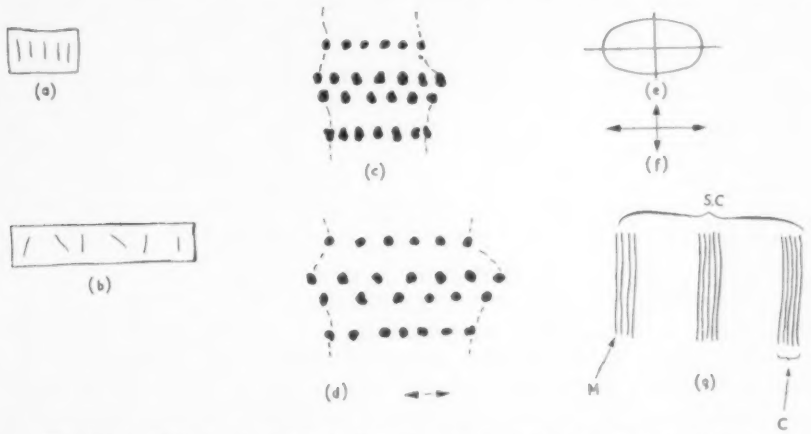


Fig. 4. (*a*) Portion of stretched film of nucleo-protein showing direction of long chain molecules. (*b*) Same film after swelling in distilled water. (*c*) Diagram of salivary gland chromosome in isotonic saline. (*d*) The same in reduced salt concentration showing lateral swelling. (*e*) Polarizability ellipsoid for a salivary gland chromosome. (*f*) Vectors showing intensity of absorption of chromosome in polarized light after staining with neutral red. (*g*) Structure of chromosomes as suggested by electron microscope studies.

M—Microfibrils 100–200 Å diameter. *Ca*—chromonema. *S.C.*—salivary gland chromosome.

and BUTLER[33]) it is found that when these films are placed in 0·14 M sodium chloride they do not swell, but in distilled water they swell markedly at right angles to their length, without longitudinal swelling. This change of shape is due, in part, to a relaxation process, such as occurs when the restraint on a sheet of stretched rubber is removed. But some birefringence remains after swelling, so that the effect appears to be due in part to a tendency for water molecules to penetrate preferentially between the somewhat rigid long chain molecules, as shown in Fig. 4*b*. A similar experiment can be carried out on unfixed cells of Drosophila or Chironomus larvae. The cells are mounted in isotonic salt (0·14 M) and the chromosomes examined in the phase contrast microscope (Fig. 5). On reducing the salt concentration of the surrounding medium, the chromosomes swell until they are quite

invisible. But it can be seen that the swelling, at right angles to the length of the long axis of the chromosome, is more marked than the swelling parallel to the long axis (2:1). The process is reversible, it being possible to restore the original appearance completely on increasing the salt concentration once more to 0·14 M. This result also suggests that the long chain molecules lie parallel to the length of the salivary gland chromosome. More information about the basic structure can be obtained by electron microscopy, although we must again be cautious about accepting results obtained on fixed cells. This is particularly true of the non-dividing nucleus, which appears to be in an extremely delicate state of colloidal equilibrium. Evidence for the existence of fibres running parallel to the length of the salivary gland chromosomes in D. Melanogaster have been obtained by PALAY and CLAUDE,[59] by YASUZUMI, ODATE and OTA,[53] and by BIRBECK and GOLDACRE.[60] YASUZUMI, ODATE and OTA found threads of ~ 100 Å diameter in the interband regions. AMBROSE, CUCKOW and GOPAL-AYENGAR[61] studied chromosomes from Chironomus salivary glands of various sizes and found threads of 100–200 Å diameter running parallel to their lengths.

In thin sections prepared according to accepted methods for the study of cytoplasmic structure, no evidence for a fibrous structure has been obtained. But it must be pointed out that BAHR[62] considers that neutral osmic solutions do not fix nucleic acid; sections also are not favourable for the demonstration of fibrous structures, unless of a high degree of order (MERCER[63]). The basic structure of salivary gland chromosomes, as suggested by optical, swelling and electron microscope studies is indicated in Fig. 4*g*.

We will now consider the macroscopic structure of these chromosomes, in order to decide how far the conclusions drawn from their study may be applied to the structure of normal cell nuclei. KOLLER[64] followed the successive stages of development of the salivary gland cells in various species of Drosophila and Chironomus. The threads at their earliest observable stage, in fixed preparations, have the appearance shown in Fig. 8*a*. Relic coils are seen in the threads, which are considered by DARLINGTON to represent the remains of the telophase coils in the last division of the cell. In each nucleus there are two chromosomes of each type, male and female, which were derived originally from sperm and egg respectively. These homologous threads gradually fuse together, as shown in Fig. 8*c*. The corresponding bands on each thread pair up and fuse. If for any reason a portion of non-homologous material is present in any given region, e.g. from reaction of the sperm with an X-ray or a chemical mutagen, fusion does not occur. At the stage of partial fusion, the remaining unpaired threads are coiled round each other (relational coiling according to DARLINGTON).

VOL
6
195

The directions of coiling are not constant, being sometimes right and sometimes left-handed. Finally the complete fusion into the cylindrical rope of the mature salivary gland chromosome occurs.

But after this fusion into the continuous rope, the salivary gland chromosome will continue to grow both in width *and* in length (BEERMANN;[65] AMBROSE, CUCKOW and GOPAL-AYENGAR[61]). No despiralization of the chromosomes occurs during this development (BEERMANN).

The mature salivary gland chromosome appears, in the light microscope to be composed after fixation of a series of strands, each strand corresponding to a chromonema (chromosome thread) of a normal somatic cell. (KOLLER's[64] polytene theory. See also PAINTER and GRIFFEN.[66]) Evidence for such a view has been obtained by SWIFT and RASCH[67] from measurements of the DNA content in comparison with normal diploid cells.

But the electron microscope studies suggest that each strand (chromonena) seen in the light microscope is multiple. We could therefore picture the salivary gland chromosome as built up in the following stages, parallel to the long axis. (Fig. 4*g*.)

(1) Nucleic acid and protein chains, 15–20 Å diameter (observed by X-rays and optical effects only).

(2) Bundles of these chains forming microfibrils 100–200 Å diameter (observed in the electron microscope).

(3) Bundles of microfibrils forming the chromonema 5000–10,000 Å diameter (observed in the light microscope), i.e. ($\frac{1}{2}$ to 1 μ).

(4) Bundles of chromonemata forming the salivary gland chromosome (up to 20 μ diameter).

It should be pointed out that METZ[68] reached the conclusion, from his studies of salivary gland chromosomes in the light microscope, that they possess an alveolar and not a fibrous structure. An attempt will be made to show later that this result is not necessarily in conflict with the polytene theory.

We should perhaps mention, at this stage, certain of the mechanical properties of chromosomes. Salivary gland chromosomes can be stretched *reversibly* both in the fixed (BUCK[69]) and in the fresh condition (D'ANGELO[70]), without any apparent loss of structure. The dimensions of the nucleus prevent extension beyond a fourfold value but after removal from the nucleus, D'ANGELO has stretched the chromosomes to ten times their original length! The interband regions stretch much more than the bands; but this does not imply that the bands are not elastic. It may be due to the fact that the interband regions, being much thinner, are subject to a greater force per unit area. Such long range elasticity certainly implies that a portion of the long chain molecules in the chromosome are present in a highly folded

state; e.g. RIS[71] has suggested that an intimate spiralization of the structure occurs in the chromomeres.

(c) *The resting nucleus of normal cells*

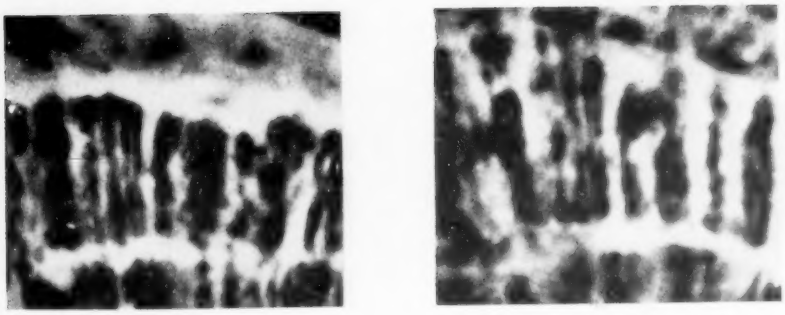
We pass on now to consider the cells of normal size in the non-dividing state. Arguing from our previous evidence on the development of the salivary cells, we might expect to find that the nucleus contained rather irregularly coiled threads (chromonemata) in which the male and female homologues were unpaired; also that the chromosomes were associated in the heterochromatic regions and that a nucleolus was present.

Unfortunately the resting nucleus does not lend itself to fixation. Most cytologists are agreed that the only regions which can be preserved are the chromocentres (regions of associated heterochromatin), and the nucleolus. The regions of heterochromatin remain clear and distinct in the resting nucleus. This is in interesting contrast to their behaviour in division where they became weakly stained (DARLINGTON and LA COUR;[73] and DARLINGTON[72]). The nucleolus can also be clearly recognized, by its staining reaction for RNA. But the problem of what happens to the chromosome threads (chromonemata) in the resting nucleus has puzzled investigators for many years.

DARLINGTON[72] has described some interesting examples of cells which have been starved of nucleic acid, a condition which enables heterochromatic regions to be recognized by their weak staining at prophase and metaphase. This occurs naturally in some sex chromosomes at meiosis and can also be produced by growing plants at low temperatures. With blood cells more general changes such as spindle defects are produced.

FUJII,[74] STROHMEYER,[75] KUWADA[76] and SINKE,[77] studied the resting nuclei of plant cells particularly in Tradescantia staminal hairs, as long ago as in 1926–1937. They were able to observe threads in the living state whose visibility depended upon the salt concentration of the medium. They suggested that the threads were always present in the nucleus even when invisible in the microscope. But until recently it had not been possible to detect threads in the nuclei of living animal cells. Improvements in optical techniques have, however, helped to clarify the situation. FELL and HUGHES[78] were able to observe the heterochromatic granules in the nuclei of tissue culture cells and also found indications of threads using the phase contrast microscope.

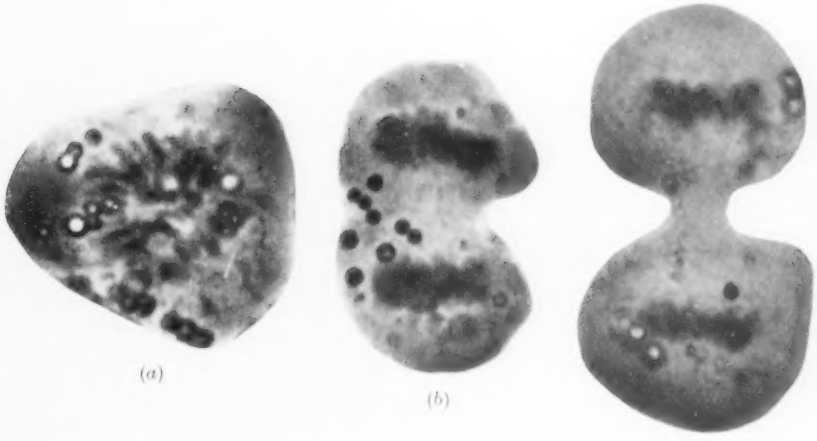
DURYEE and DOHERTY[79] have been able to show that conditions favourable to the precipitation of nucleo-protein lead to the appearance of rod-like structures of about the diploid number in the resting nuclei of thecal cells in ovarian eggs. AMBROSE has shown that the difficulty in seeing the threads in the undisturbed resting nucleus was simply due



(*a*) (*b*)

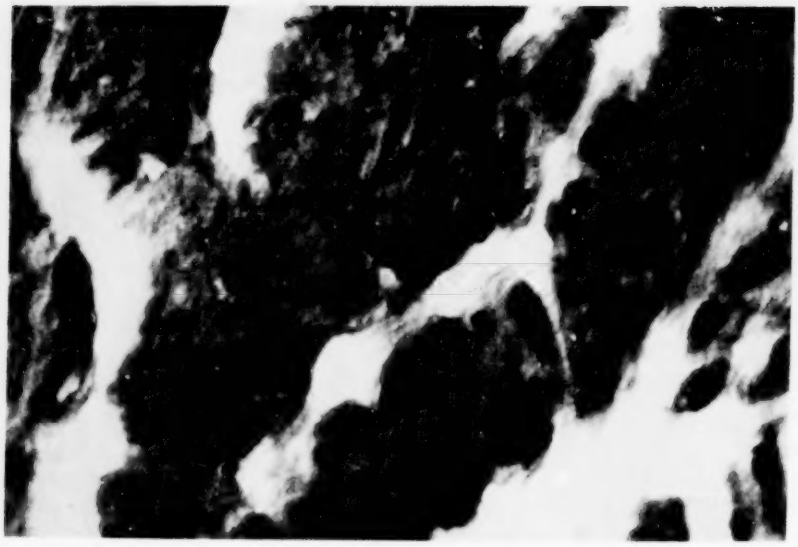
Fig. 5. (*a*) Salivary gland chromosome in living cell in isotonic saline (axis of chromosome horizontal). (*b*) The same chromosome in reduced salt concentration showing lateral swelling. (Compare with Fig. 4*c*, *d*.)

VOL.
6
956



(*a*) (*b*) (*c*)

Fig. 6. Chromosomes in living cells in division. (*a*) Polar view of metaphase in Erlich ascites cell. (*b*) Anaphase. (*c*) Telophase.



(*a*)

(*b*)

Fig. 7 (*a*) and (*b*). Electron micrograph of human pachytene chromosomes. Negative contrast with chromosome light on dark field.

VOL.
6
956

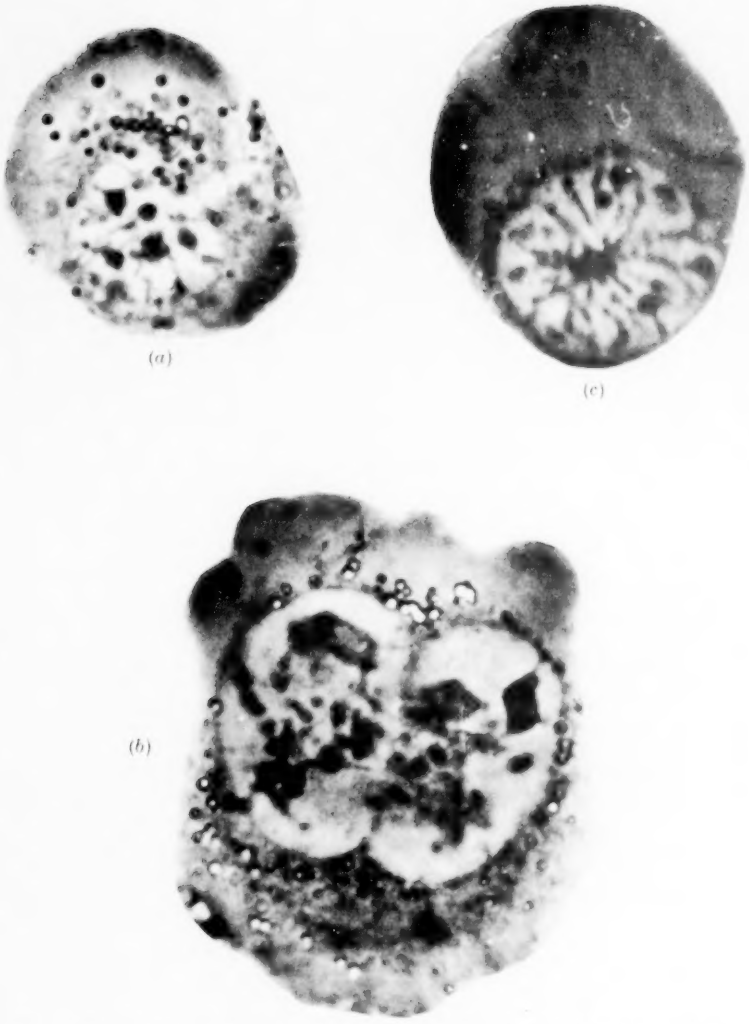


Fig. 8. Interference microscope photographs of the resting nuclei of living cells. (*a*) Erlich ascites tumour cell showing heterochromatic masses and thin chromosome threads in nucleus. (*b*) Giant cell of same tumour showing grossly enlarged and vacuolated masses also fine threads. (*c*) Late resting stage in Neuroblast of Locusta showing chromosome threads attached to dense chromocentre.

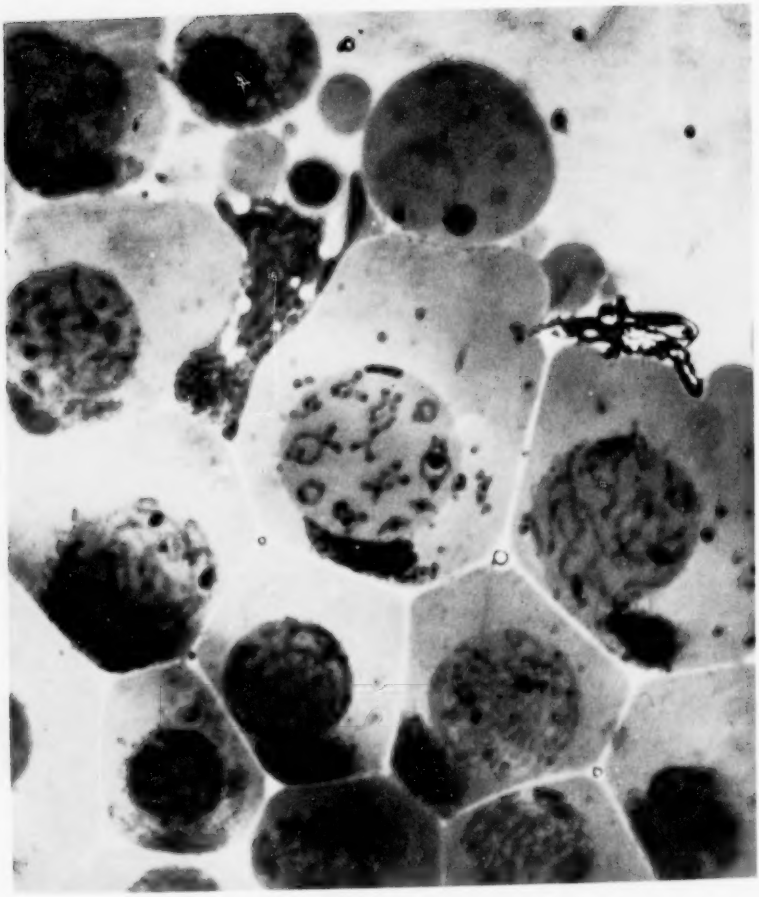


Fig. 9. Meiosis in fresh cells of Locusta males, showing cells in leptotene, bouquet, pachytene and diakinesis stages. (Interference microscope photograph.)

VOL
6
195

to lack of a sufficiently sensitive method of detecting small differences of refractive index. With the interference microscope threads have been seen in the following resting nuclei; Spermatogonia, spermatocytes and neuroblasts (with GOPAL-AYENGAR) of Locusta. Chick heart fibroblasts, Erlich Ascites tumour cells and Krebs Ascites tumour cells (Fig. 9*e*). In a number of cases the threads can be seen to be attached to the inflated heterochromatic regions where they are sometimes

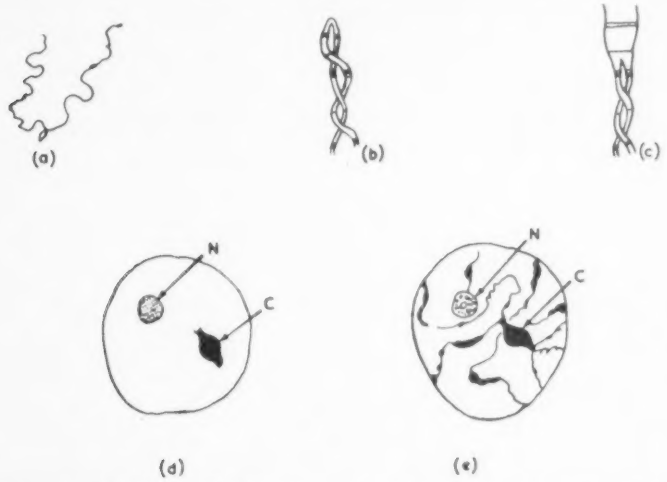


Fig. 10. (*a*) Salivary gland chromosome in very early stage showing relic coils (KOLLER). (*b*) Later stage showing relational coiling of male and female chromosomes. (*c*) Stage showing partial fusion of male and female portions at homologous points. (*d*) The resting nucleus of normal cells as seen in ordinary microscope. *N*—nucleolus, *C*—chromocentre. (*e*) The structure of the resting nucleus as indicated by interference microscopy.

associated in chromocentres. In the author's view there is reason to believe that the resting nucleus of normal cells is similar on a smaller scale to the salivary gland cells, except that there may be a relatively higher degree of development of heterochromatin in some cases in the small cells. This is particularly marked in tumour cells (Fig. 8).

VI. CHROMOSOMES IN DEVELOPMENT

It may be of interest, at this stage, to mention some promising work which has been carried out during recent years on the changes in chromosome appearance which occur when it is thought that they are engaged in some kind of secretory process; i.e. when the genetic material may be at work in deciding the development of the organism. BEERMANN[80] has found that the giant chromosomes observed in different larval tissues are not identical, e.g. although the number and sequence of the bands is the same in the salivary glands and in the

Malphigian layers, the dimensions of the bands differ in the two tissues, suggesting that the relative degrees of activity of genetic elements are not the same. An extreme case of this effect is the Balbianni ring structure shown diagrammatically in Fig. 11*b* and also studied by BEERMAN and BOHR.[81] These are regions in which a fan-like development of material takes places from a single chromosome band as shown in the electron microscope. MECHELKE[82] has shown in a most interesting study that the degree of development and number of Balbianni

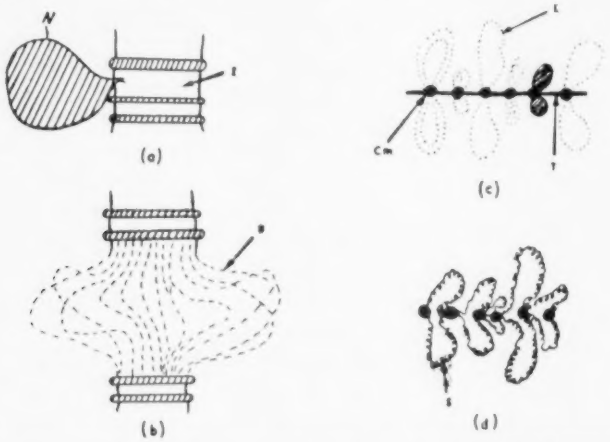


Fig. 11. Chromosomes in development. (*a*) The development of a nucleolar structure from a single interband region in an S.G. chromosome (after BEERMANN). (*b*) Balbianni ring development in S.G. chromosome (after BEERMANN). (*c*) Lamp brush chromosome showing development of lateral loops (DURYEE, CALLAN). (*d*) Appearance of spiral form in lateral loops of lamp brush chromosome (after RIS).

rings changes with the age of the larva in the salivary glands and they can be correlated with the reversible appearance and disappearance of a brown secretion by the gland.

An extreme case of this hyperactivity of chromosomes is perhaps shown by the well-known lamp brush chromosomes of Amphibian oocytes. In this case there is a rapid synthesis of cytoplasmic storage material taking place in preparation for the development of the egg. In these nuclei, lateral loops of various shapes and sizes are seen attached to the chromosomes (Fig. 11*c*). According to DURYEE,[83] TOMLIN and CALLAN[84] and GALL,[85] there is a central thread (chromonema) carrying the chromosomes with the loops formed as lateral attachments (Fig. 11*c*). But according to RIS,[86] who has taken stereoscopic electron microscope pictures, the loops are continuous with the main thread and are in various degrees of spiralization (Fig. 11*d*).

But whichever view is taken about the structure, it is clear that the lateral loops represent some secretory product analogous to the Balbianni rings in the salivary gland chromosomes. It may be mentioned that DARLINGTON[87] suggested that the growth of the salivary gland chromosomes themselves was due to the appearance of gene products.

VII. CHROMOSOMES IN DIVIDING CELLS

(a) *The synthesis of chromosomes*

We have considered so far the composition and structure of chromosomes within the nuclei of non-dividing cells. An essential requirement for cell division is that a duplication of the structural elements in the cell should take place. In the case of chromosomes this process has been followed very beautifully by studying the deoxyribose nucleic acid content of the cell nuclei. BOIVIN, VENDRELY and VENDRELY,[88] and VENDRELY and VENDRELY[89] showed that, in non-dividing tissue cells, the amount of DNA per nucleus was constant, for a diploid tissue, that it was halved in the sperm which contained a haploid number of chromosomes and that it was doubled in tetraploid cells. Similar results have been obtained by MIRSKY and RIS,[90] POLLISTER, SWIFT and ALFERT,[91] THOMSON and FRAZER[92] and by HARRISON.[93] In the case of dividing cells, WALKER and YATES[94] have shown they have twice the diploid quantity of DNA immediately after telophase, but this is increased to 4 $\times$ the diploid amount before prophase. There are one or two apparent exceptions to the principle of DNA constancy per chromosome, e.g. MARSHAK and MARSHAK[95] found that little or no DNA could be detected in the pronucleus of Arabacia eggs. But it must be pointed out that the analysis of the egg nucleus is extraordinarily difficult, in view of the great volume of cytoplasm in relation to the nuclear volume. Other investigators have found abnormally high values for the DNA content of eggs. (SCHMIDT, HECHT and THANNHAUSER;[96] VENDRELY, VENDRELY;[97] and JØRGENSEN and ZEUTHEN.[98]) But should the result of MARSHAK and MARSHAK be confirmed with other types of egg, an entirely new approach to the DNA problem will certainly be required.

In so far as the synthesis of nuclear protein is concerned there are as yet few experimental observations available. It may be mentioned that VENDRELY and VENDRELY[99] have found a quantitative relationship between the nuclear histone and the DNA, suggesting that some mutual relationships are involved between nucleic acid and basic protein during the synthetic process.

Molecular mechanisms for the synthetic processes have been proposed by numbers of investigators. It may be of interest to consider

the proposal of CRICK and WATSON that the two chains in their helix are identical but complementary so that synthesis occurs by the duplication of one of the chains. There are various biological objections to this hypothesis, namely:

(1) The length of the two chain model on the present evidence for the molecular weight of DNA exceeds the known length of the chromosome threads.

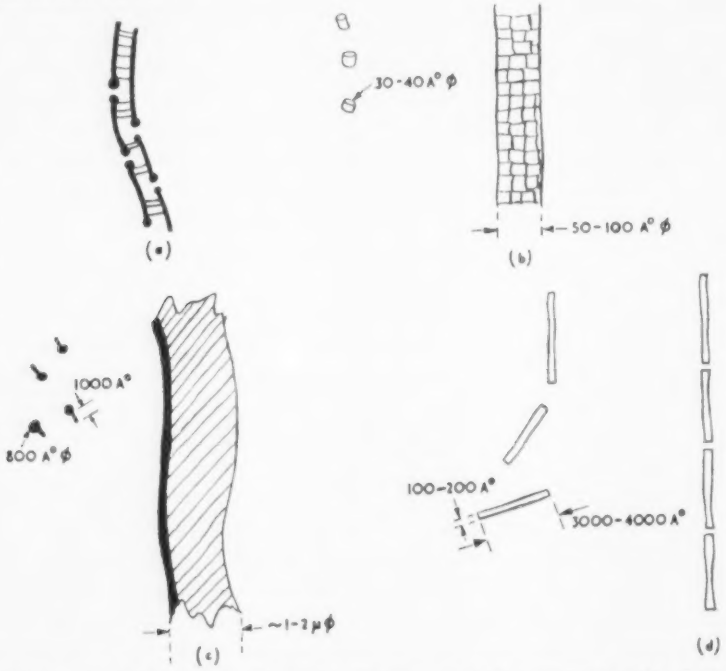


Fig. 12. (*a*) Model for nucleic acid (DEKKER and SCHACHMAN) with short lengths of chain held together by hydrogen bonds. (*b*) Free insulin molecules and fibre of fibrous form. (*c*) Free T_2 phage virus particles and fibre formed from particles. (*d*) Possible structural units of chromosome and fibre formed from aggregation of particles by secondary forces of attraction.

(2) The strong adhesion between the two chains produced by a large number of hydrogen bonds.

(3) The difficulty in unwinding the helices which are interlocked (orthospiral).

In order to avoid these difficulties, DEKKER and SCHACHMAN[100] have proposed that the nucleic acid molecule is in fact composed of quite short chains of 30–100 nucleotides, on the basis of end group analysis and molecular weight determinations of mildly heated or acid treated material. The model is illustrated in Fig. 12*a*, the short nucleic acid chains are in this case linked together by hydrogen bonds to form

the large molecular weight structure. This model overcomes certain difficulties associated with the model of CRICK and WATSON.

But the synthesis of chromosomes is not simply a problem involving the synthesis of nucleic acid and protein chains (stage I). Quite large scale structures must be built up. Evidence has been obtained by AMBROSE, CUCKOW and GOPAL-AYENGAR, that the micro-fibrils of 100–200 Å diameter are seen both in large and small Chironomous salivary glands and in human pachytene (meiotic) chromosomes (Fig. 7). Threads of about the same dimensions have also been observed by Tomlin and Callan in lamp-brush chromosomes. These results suggest that Stage II in the synthetic process may be the formation of

Fig. 13. (*a*) Salivary gland chromosome showing partial damage by chemical mutagens (FAHMY and BIRD). (*b*) Partial damage in endosperm anaphase bridge (LA COUR and RUTISHAUSER).

micro-fibrils of these dimensions. Independent biological evidence for a multiple stranded structure of the chromonema (the thread seen in the light microscope) is available; FAHMY and BIRD[101] have found that treatment of the sperm of Drosophila males (at a stage when only one chromonema is present) can lead to partial damage observed in a portion of one-half of the paired salivary gland chromosomes of the offspring (Fig. 13*a*). Such a condition could only arise if partial damage of the chromonema had occurred in the sperm. LA COUR and RUTISHAUSER[102] have shown that, in endosperm, the bridges which sometimes form at anaphase after X-irradiation are due to partial damage and exchange between subchromatids (Fig. 13*b*). DARLINGTON[72] has described instances in meiosis where only a partial crossing over has occurred, which can also be explained by a multiple stranded structure.

These observations do not imply that the genetic elements of the chromosome are themselves multiple but merely that they are physically composed of multiple strands and that chromosome synthesis involves their duplication. Evidence for a polygenic character of the hereditary material must rest also on genetic experiments.

The question remains as to whether the microfibrils should be considered to be composed of a continuous nucleic acid or protein chain. As already mentioned, the evidence of the Caspersson School and the enzymatic studies, suggest that the nucleic acid is not deposited continuously along the chromonema, but is concentrated in the chromomeres. Also the evidence of DEKKER and SCHACHMAN suggests that the actual molecule may not be a continuous fibre. The proteins have a molecular weight too low to enable them to form a continuous chain along the length of the chromosome. So the alternative possibility is that the microfibrils are formed by the aggregation of small particles. It is now considered that a number of different types of biological fibre are formed by such a process (Fig. 12). Insulin molecules will aggregate to produce uniform fibres of ~ 60 Å minimum diameter by hydrogen bond formation between the molecules (WAUGH;[103] LENS;[104] AMBROSE and ELLIOTT;[105] FARRANT and MERCER[106]). Wool keratin appears to be produced by such a process (FARRANT, REES and MERCER[107]); RANDALL[108] considers that collagen fibres are formed in this way. As already mentioned in Section IV, T_2 phage virus particles will produce fibres. In the case of insulin fibres, the hydrogen bonds between the molecules can be broken by the phenolic reagents (compounds which also transform the *inter*-chain bonded β form of protein chain into the *intra*-chain bonded folded form). The effect of these compounds on fresh salivary gland chromosomes within the cell was tested by AMBROSE and GOPAL-AYENGAR.[109] With phenol at 0·5 M concentration the chromosomes disperse rapidly with slight longitudinal swelling. There is no lateral swelling. (This is in contrast to the effect of reducing the salt concentration which causes *lateral* swelling.) This result suggests that hydrogen bonds may be partly responsible for the longitudinal cohesion of the chromosome thread, in contrast to the lateral forces which appear to be of an electrostatic nature. AMBROSE and GOPEL-AYENGAR suggested that the chromosome thread might be formed by the aggregation of nucleoprotein particles, these particles being held together by secondary forces of cohesion, possibly hydrogen bonds. This does not imply that the polypeptide chains do not represent structural units within the particles, as suggested by the enzymatic studies, but that points of weakness, possibly between genetic elements, occur along the chromosome thread, these attachments being maintained by secondary forces. An independent biological experiment appears to support this view. LEVAN and TJIO,[110] and TJIO[111] studied the effect of very low concentrations of phenolic compounds (0·01–0·0005 M) on growing root tips. Similar experiments have been carried out by other investigators using nitrogen mustards and other chemical mutagens (DARLINGTON and KOLLER[112]). But whereas the chemical mutagens (which are

thought to combine with genetic material) and x-rays produce permanent alterations in chromosome structure (fragmentation, etc.), LEVEN and TJIO found that fragmentation produced by phenolic compounds was not permanent. The effects of these compounds were of a temporary character. Such effects could be understood if the compounds were merely breaking the reversible adhesions along the chromosome thread.

MAZIA[114] has reached a similar conclusion about chromosome structure, but he considers that the bonds holding the particles together along the length of the thread are due to calcium ions (—] Ca^{++} [—) (see also BERNSTEIN and MAZIA[113]). Evidence in favour of this view is given by the fact that the chromosomes can be dispersed by the combination of salt solutions and citrate, which absorbs calcium ions. MAZIA has in fact been able to produce particles of length 4000 Å from chromosomes, which can be seen in the electron microscope. Incidentally, evidence for nucleo-protein particles of minimum length 3000 Å in extracted material has been obtained by STEINER.[115]

The sensitivity of nucleo-protein to the presence of bivalent cations has already been described (Section III), and here again an interesting biological study of the effect of bivalent cations has been made. STEFFENSEN[116] has found that, with Tradescantia Paludosa, plants grown on HAAGLAND's magnesium deficient medium show a nine-fold increase in the number of chromosome fragments at meiosis.

In the author's opinion there is therefore good evidence in favour of the view that chromosome threads (chromonemata) are produced by the aggregation of particles of nucleo-protein, in which the adhesive forces are due in part to hydrogen bonds and in part to linkage by bivalent cations. It was mentioned in Section II, that there are some indications that the basic proteins, when combined with nucleic acid are transformed from the soluble non-interacting *folded* configuration into the β form, which associates with other molecules to produce a precipitate or a gel-like structure. BUTLER has recently suggested that the two types of histone (*lysine* rich and *arginine* rich fractions) might play different roles in the formation of chromosome structure, one being connected with lateral and the other with longitudinal linkages of the DNA molecules.

The model for chromosome structure which is based on the aggregation of particles may help to reconcile some of the rather contradictory views about the fibrous structure which have been mentioned earlier in this review. A fibrous material formed in this way could be changed readily by fixation into the network or alveolar form of METZ, and could in fact achieve such a state under certain biological conditions. It is also possible that the continuous thread observed by DURYEE, CALLAN and TOMLIN and GALL in the lamp brush chromosomes could readily

break up to give the more open structure held together only by lateral loops, as seen by Ris in the electron microscope.

(b) *Chromosomes in division*

It remains but to consider briefly the behaviour of chromosomes in dividing cells. This problem has been investigated systematically by cytologists for many years. We may examine, for example, the classic work of Darlington[117] on the internal and external mechanics of the chromosomes in Mitosis and Meiosis.

The most significant observable change which takes place in dividing cells involves a process of spiralization. The spiralization was beautifully demonstrated by Kuwada and Nakamura[118] by pretreating the chromosomes before fixation with hypertonic salts or with calcium ions, in the case of Tradescantia meiotic cells. In mitosis there is one spiral (minor spiral, Darlington) whereas in meiosis there are two (minor and major, Darlington).

From purely general considerations there are three possible ways of producing the spiral. Darlington considers:

(1) Rotation of one end according to the number of turns in the helix.

(2) Compensation of each turn by an internal twist of the thread in the opposite sense.

(3) Winding successive short segments of the thread in opposite directions.

Darlington considers that (1) is excluded because ring chromosomes spiralize as readily as those with open ends (Navashim[119]); also long and short chromosomes spiralize at equal rates. He considers that (3) is excluded because the spiralized appearance of the chromosomes is uniform over considerable lengths, e.g. along the whole arm of a chromosome with central centromere. Darlington therefore refers to a compensating *molecular* spiral which is formed in the opposite sense to the direction of spiralization of the visible minor and major spirals. He considers that the formation of the molecular spiral represents the first stage in the contraction of the chromosome threads as the cell comes towards division at early prophase. The visible twists which are seen in the threads at this stage are due to relic spirals left over from the previous division (telophase) (Fig. 14).

As the cell comes towards prophase in mitosis, the first visible manifestation of the duplication process described in the previous section appears. It is seen that the chromosome is split longitudinally; each daughter chromatid is spiralized and shortens continuously as metaphase is approached. Darlington considers that the two threads, before splitting takes place, are arranged in an anorthospiral (Fig. 14*b*). In this type of spiral the two portions are not interlocked and can separate laterally without unwinding. This type of spiral must be

VOL
6
195

contrasted with that shown in the CRICK and WATSON model for nucleic acid in which the two threads are wound round each other and interlocked (orthospiral, Fig. 14*a*)

MANTON[120] has reviewed the more recent studies of chromosome spiralization and considers that their direction of spiralization is not constant. It may even differ in the two daughter chromatids

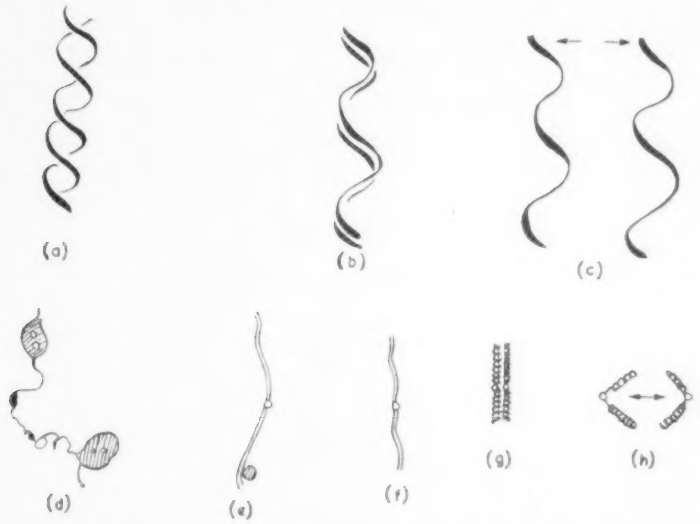


Fig. 14. (*a*) Orthospiral. (*b*) and (*c*) Anorthospiral. (*d*) Chromosome in resting nucleus. (*e*) Chromosome in early prophase. (*f*) Chromosome in late prophase. (*g*) Chromosome in metaphase. (*h*) Chromosome in anaphase.

(MANTON,[123] MANTON and SMILES[121]). She finds that ortho spirals almost certainly do not occur in the chromosomes.

In addition to the process of spiralization, a number of other changes take place, as the cell passes from the resting stage into division. These are summarized in Fig. 14 and in Table II. Very few of these changes can at present be described in physio-chemical terms. As DARLINGTON has pointed out, the electrostatic theory of chromosome movement of LILLIE[122] remains the best. But we should now begin to think perhaps in terms of interactions between polyelectrolytes (long chain molecules with charged groups attached to them) rather than in terms of simple electrostatic charges.

A specific arrangement of changed groups along the chromosome thread may help to explain secondary pairing, for example. In this process, homologous chromosomes are seen to lie parallel to each other on the metaphase plate in certain cases in mitosis and meiosis (DARLINGTON[123]).

TABLE II

Chromosomes in dividing cells

Property observed	*Interphase*	*Prophase*	*Metaphase*	*Anaphase*	*Telophase*	*Interphase*
Spiralization of chromosome	Extended. Some relic coils	Contraction of molecular spiral	Contraction of minor spiral. Separation of chromatids	Separation of centromeres	Highly condensed	Rapid relaxation of molecular spiral. Partial relaxation of minor spiral
Concentration of chromosome material	15–18%	Increasing	Max. 33%		Decreasing	12–15%
Ash content		Increasing	Some calcium?		Decreasing	
Heterochromatin	Well developed. Condensed	Loss of nucleic acid	Similar to euchromatin. (Can be starved of N.A.)			Remains condensed
Nucleolus	Continuous growth	Almost lost	Absent		Absent	Develops from nuclear nucleolar organizer
Nuclear membrane	Thick and distinct	Disappears				Reforms

VOL. 6
195

We might go on to consider the more detailed behaviour of chromosomes in meiosis, e.g. the process of chromosome pairing in meiosis. In this process, the various sets of chromosomes pair up in homologous regions as they do in salivary gland chromosomes. During this stage a four stranded appearance of the structure develops, two strands being derived from each homologous chromosome. But interchange between parts of these strands can take place, leading to a reassortment of the genetic elements between the homologous threads (Fig. 15). The chiasmata which result, when the separation of the chromosomes takes place during the first meiotic division, can be seen in the interference microscope photograph of a living cell shown in Fig. 10. These interchanges can be correlated with the crossing effects observed by

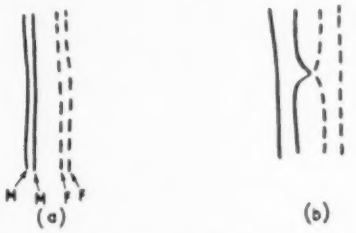


Fig. 15. Chiasma formation in meiosis.
M—Male chromonemata. *F*—Female chromonemata.

direct genetic experiments. The interchanges take place at the four stranded stage; the detailed mechanics in relation to the tensions and coiling processes have been described by DARLINGTON[121] who has also considered the conditions which give rise to the difference between meiotic and mitotic division.[123] In all such problems it must be remembered that the chromosomes cannot be considered in isolation, but are present in a biological system in dynamic equilibrium, where structures are preserved by a continuous flow of energy and building materials through the cell. Alterations in relative *rates* of synthesis can effect the final result in a marked degree; many changes of chromosome composition, seen in abnormal cells, can in fact be traced to such causes.

But these are matters which can provide problems for physical chemists for many years to come. This review will have achieved its purpose if it has given an indication of the various physical methods which are being applied to the study of that beautiful but elusive entity, the chromosome.

The author wishes to thank Professor A. HADDOW for his interest and encouragement in the investigations of chromosome structure described in this chapter. He is grateful to Professor C. D. DARLINGTON, F.R.S., for helpful comments and discussions, and to Professor J. A. V.

BUTLER and Professor P. C. KOLLER. This investigation has been supported by grants to the Chester Beatty Research Institute (Institute of Cancer Research: Royal Cancer Hospital) from the British Empire Cancer Campaign, Jane Coffin Childs Memorial Fund for Medical Research, the Anna Fuller Fund, and the National Cancer Institute of the National Institutes of Health, U.S. Public Health Service.

REFERENCES

(1) DAVISON, P. F., CONWAY, B. E. and BUTLER, J. A. V.; *Progress in Biophysics and Biophysical Chemistry* 4 (1954) 148
(2) FELIX, K., FISCHER, H. and KREKELS, A.; This volume, p. 1
(3) DAVISON, P. F., BUTLER, J. A. V., JAMES, D. W. F. and SHOOTER, K. V.; *Biochim. et Biophys. Acta*
(4) DALY, M. and MIRSKY, A. E.; *J. Gen. Physiol.* 38 (1955) 405
(5) CRAMPTON, C. F. and MOORE, S.; *J. Biol. Chem.* (In Press)
(6) STEDMAN, E. and STEDMAN, E.; *Biochem. J.* 38 (1944) xxvi
(7) CROFT, H. J., MAURITZEN, C. M. and STEDMAN, E.; *Nature, Lond.* 174 (1954) 581
(8) STEDMAN, E. and STEDMAN, E.; *Nature, Lond.* 152 (1943) 267
(9) CRICK, F. H. C. and WATSON, J. D.; *Proc. Roy. Soc.* A 223 (1954) 80
(10) CHARGAFF, E., VISCHER, E. and ZAMENHOF, S.; *J. Biol. Chem.* 173 (1948) 327
(11) CHARGAFF, E., CRAMPTON, C. F. and LIPSCHITZ, R.; *Nature, Lond.* 172 (1953) 289
(12) LUCY, J. A. and BUTLER, J. A. V.; *Biophys. Biochim. Acta* 16 (1955) 437
(13) VINCENT, W. S.; *Proc. U.S. Nat. Acad. Sci.* 38 (1952) 139
(14) ELSON, D. and CHARGAFF, E.; *Nature, Lond.* 173 (1954) 1037
(15) POLLISTER, A. W. and LEUCHTENBERGER, C.; *Nature, Lond.* 163 (1949) 360
(16) ASTBURY, W. T. and STREET, A.; *Phil. Trans.* A 230 (1931) 79
(17) ASTBURY, W. T. and WOODS, H. J.; *Phil. Trans.* A 232 (1933) 333
(18) AMBROSE, E. J. and HANBY, W. E.; *Nature, Lond.* 163 (1949) 483
(19) AMBROSE, E. J. and ELLIOTT, A.; *Proc. Roy. Soc.* A 205 (1951) 47; 206 (1951) 206; 208 (1951) 75
(20) BAMFORD, C. H., HANBY, W. E. and HAPPEY, F.; *Proc. Roy. Soc.* A 205 (1951) 30; 206 (1951) 407
(21) PAULING, L., COREY, R. B. and BRANSON, H. R.; *Proc. Nat. Acad. Sci. Wash.* 37 (1951) 205
(22) PAULING, L. and COREY, R. B.; *Proc. Nat. Acad. Sci. Wash.* 37 (1951) 235, 241, 256, 261
(23) PERUTZ, M. F.; *Nature, Lond.* 167 (1951) 1053
(24) BAMFORD, C. H., BROWN, L., ELLIOTT, A., HANBY, W. E. and TROTTER, I. F.; *Proc. Roy. Soc.* B 141 (1953) 49
(25) WALEY, S. S. and WATSON, J.; *Biochem. J.* 55 (1953) 328
(26) ELLIOTT, A.; *Proc. Roy. Soc.* A 226 (1954) 408
(27) SADRON, C.; *Progress in Biophysics*, 3 (1953)
(28) POUYET, J.; *J. Chim. Phys.* 48 (1951) 90
(29) CONWAY, B. E. and BUTLER, J. A. V.; *J. Polymer. Sci.* 11 (1953) 277
(30) WILKINS, M. H. F. and RANDALL, J. T.; *Biochim. et Biophys. Acta* 10 (1953) 1992
(31) RILEY, D. P. and OSTER, G.; *Discuss. Faraday Soc.* 11 (1951) 107
(32) FRASER, M. J. and FRASER, R. D. B.; *Nature, Lond.* 167 (1951) 759

VOL.
6
1956

REFERENCES

(33) AMBROSE, E. J. and BUTLER, J. A. V.; *Discuss. Faraday Soc.* 13 (1953) 261
(34) WILKINS, M. H. F.; *Pubbl. Staz. Zool. Napoli.* 32 (1951) 104
(35) WHITE, J. C. and ELMES, P. C.; *Nature, Lond.* 169 (1952) 151
(36) AMBROSE, E. J. and GOPAL-AYENGAR, A. R.; *Nature, Lond.* 169 (1952) 652
(37) ASTBURY, W. T.; *Symp. Soc. Exp. Biol.* 1, Nucleic Acid (1947), Camb. Univ. Press 66
(38) FRASER, R. D. B.; *Discuss. Faraday Soc.* 13 (1953) 284
(39) AMBROSE, E. J.; (in the course of publication)
(40) ELLIOTT, A. and AMBROSE, E. J.; *Discuss. Faraday Soc.* 9 (1950) 251
(41) SHOOTER, K. V., DAVISON, P. F. and BUTLER, J. A. V.; *Biochim. Biophys. Acta* 13 (1954) 192
(42) SCHMIDT, W. J.; *Chromosoma* 2, (1) (1941) 86
(43) BALBIANNI, E. G.; *Zool. Anz.* 4 (1881), 637, 662
(44) ENGSTRÖM, A. and RUCH, F.; *Proc. Nat. Acad. Sci. Wash.* 37 (1951) 459
(45) CASPERSSON, T.; *Chromosoma* 1 (1940) 526; *Cell Growth and Function*, W. W. Norton and Co. Publ., 1950
(46) SCHULTZ, J.; *Proc. VII Intern. Genet. Congr.*, 1939, p. 257
(47) GOLDSCHMIDT, R. B.; *Proc. VIII Intern. Genet. Congr.*, suppl. to *Hereditas* (1949) 244
(48) PATTERSON, E. K., LANG, H. M., DACKERMAN, M. G. and SCHULTZ, J.; *Exp. Cell. Res.* 6 (1954) 181
(49) SCHULTZ, J.; *Cold Spring Harbor Symp. Quant. Biol.* 12 (1947) 179
(50) BRACHET, J.; *Embryologie Chimique.* Maison et Cie, Paris, 1944
(51) DANIELLI, J. F.; *Cold Spring Harbor Symp. Quant. Biol.* 14 (1950) 32
(52) MAZIA, D. and JAEGER, L.; *Proc. Nat. Acad. Sci. Wash.* 25 (1939) 459
(53) YASUZUMI, G., ODATE, Z. and OTA, Y.; *Cytologia* 16 (1951) 3
(54) MCDONOUGH, E. S., ROWAN, R. and MOHN, H.; *J. Hereditas* 43 (1952) 3
(55) KAUFMANN, B. P., GAY, H. and MCDONALD, M.; *Genetics* 35 (1950) 117
(56) KAUFMANN, B. P.; *Exptl. Cell Research* 4 (1953) 408
(57) PFEIFFER, H. H.; *Chromosoma* 2 (1) (1941) 77
(58) RUCH, F.; *Exp. Cell Res.* 2 (1951) 680
(59) PALAY, S. L. and CLAUDE, A.; *J. Exp. Med.* 89 (1949) 431
(60) BIRBECK, M. S. C. and GOLDACRE, R. J.; Private communication
(61) AMBROSE, E. J., CUCKOW, F. W. and GOPAL-AYENGAR, A. R.; *Int. Cyt. Cong. Leiden* (1954). Publ. under I.U.B.S.
(62) BAHR, C. F.; *Exp. Cell Res.* 7 (1954) 457
(63) MERCER, H.; Private communication
(64) KOLLER, P. C.; *Proc. Roy. Soc.* B 118 (1935) 371
(65) BEERMANN, W.; *Chromosoma* 5 (1952) 139
(66) PAINTER, T. S. and GRIFFEN, A. B.; *Genetics* 22 (1937) 612
(67) SWIFT, H. and RASCH, E. M.; *J. Histochem. Cytochem.* 2 (1954) 456
(68) METZ, C. W.; *Cold Spring Harbor Symp. Quant. Biol.* 9 (1941) 23
(69) BUCK, J. B.; *J. Heredity* 33 (1942) 3
(70) D'ANGELO; *Biol. Bull.* 90 (1946) 71
(71) RIS, H. and CROUSE, H.; *Proc. Nat. Acad. Sci. Wash.* 31 (1945) 321
(72) DARLINGTON, C. D.; *Proc. 8th Int. Congr. Genet.* Suppl. *Heriditas* (1949) 189
DARLINGTON, C. D.; *Symp. Soc. Exp. Biol.* 1, Nucleic Acid, Camb. Univ. Press, 1947
(73) DARLINGTON, C. D. and LA COUR, L. F.; *J. Genet.* 40 (1940) 185
(74) FUJII, K.; *Rept. Jap. Assocn. Sci. (Kyato Meeting) Tôyôgakugeizassi* 43 (1926)
(75) STROHMEYER, S.; *Z. Exp. Zytologie Planta* 24 (1935)
(76) KUWADA, Y.; *Cytologia, Fujii Jub. No.* (1937), 389, 402

(77) SINKE, N.; *Mem. Coll. Sci. Kyato Imp. Univ. S.B.* 15 (1939)
(78) FELL, H. B. and HUGHES, A. F.; *Quart. J. Micr. Soc.* 90 (1949) 355
(79) DURYEE, W. R. and DOHERTY, J. K.; *Ann. N.Y. Acad. Sci.* 58, 1210
(80) BEERMAN, W.; *Chromosoma* 5 (1952) 139
(81) BEERMAN, W. and BOHR, G. F.; *Exp. Cell Res.* 6 (1954) 195
(82) MECHELKE, F.; *Chromosoma* 5 (1953) 511
(83) DURYEE, W. R.; *Ann. N.Y. Acad. Sci.* 50 (1950) 920
(84) TOMLIN, S. G. and CALLAN, H. G.; *Quart. J. Microscop. Sci.* 92 (1951) 221
(85) GALL, J. C.; *Expt. Cell. Res. Suppl.* 2 (1952) 95
(86) RIS, H.; *Records Genet. Soc. Amer.* 21 (1952) 63
(87) DARLINGTON, C. D.; *Proc. 8th Int. Genet. Cong., Hereditas Suppl.* (1949) 189
(88) BOIVIN, A., VENDRELY, R. and VENDRELY, C.; *C.R. Acad. Sci.* (*Paris*) 226 (1948) 1061
(89) VENDRELY, R. and VENDRELY, C.; *Experientia* 4 (1948) 434
(90) MIRSKY, A. E. and RIS, H.; *Nature, Lond.* 163 (1949) 666
(91) POLLISTER, A. W., SWIFT, H. and ALFERT, M.; *J. Cell. Comp. Physiol.* 38 (1951) 101
(92) THOMSON, R. Y. and FRAZER, S. C.; *Exptl. Cell. Res.* 6 (1954) 367
(93) HARRISON, M. F.; *Nature, Lond.* 168 (1951) 248
(94) WALKER, J. M. B. and YATES, N. B.; *Proc. Roy. Soc. Lond.* B 140 (1952) 274
(95) MARSHAK, A. and MARSHAK, C.; *Nature, Lond.* 174 (1954) 919
(96) SCHMIDT, G., HECHT, L. and THANNHAUSER, S. J.; *J. Gen. Physiol.* 31 (1948) 203
(97) VENDRELY, R. and VENDRELY, C.; *C.R. Acad. Sci.* (*Paris*) 235 (1952) 444
(98) HOFF-JØRGENSEN, E. and ZEUTHEN, E.; *Nature, Lond.* 169 (1952) 245
(99) VENDRELY, R. and VENDRELY, C.; *C.R. Acad. Sci.* (*Paris*) 235 (1952) 395
(100) DEKKER, C. A. and SCHACHMAN, H. K.; *Proc. Nat. Acad. Sci. Wash.* 40 (1954) 894
(101) FAHMY, O. G. and BIRD, M. J.; *Proc. Roy. Soc.* B 140 (1953) 556
(102) LA COUR, L. F. and RUTISHAUSER, A.; *Nature, Lond.* 72 (1953) 501
(102a) DARLINGTON, C. D.; *Pubbl. Staez. Zool. de Napoli* 22 (1950) (Suppl.) 22
(103) WAUGH, D. F.; *J. Amer. Chem. Soc.* 68 (1946) 247
(104) LENS, J.; *J. Biol. Chem.* 169 (1947) 313
(105) AMBROSE, E. J. and ELLIOTT, A.; *Proc. Roy. Soc.* A 208 (1951) 75
(106) FARRANT, J. L. and MERCER, E. H.; *Biochim. Biophys. Acta* 8 (1952) 355
(107) FARRANT, J. L., REES, A. L. and MERCER, E. H.; *Nature, Lond.* 159 (1947) 535
(108) RANDALL, J. T. (Ed.); *The Nature and Structure of Collagen*, Butterworths 1953, p. 232
(109) AMBROSE, E. J. and GOPAL-AYENGAR, A. R.; Suppl. to *Heredity* 6 (1952) 277
(110) LEVAN, A. and TJIO, J. H.; *Hereditas* 34 (1948) 250; *Annee Biol.* 52 (1948) 192
(111) TJIO, J. H.; *Anales estac. exptl. Qula Dei* (*Zaragoza*) 2 (1951) 187
(112) DARLINGTON, C. D. and KOLLER, P. C.; *Heredity* 1 (1947) 187
(113) BERNSTEIN, M. H. and MAZIA, D.; *Biochim. Biophys. Acta* 10 (1953) 600; 11 (1953) 59
(114) MAZIA, D.; *Proc. Nat. Acad. Sci. Wash.* 40 (1954) 521
(115) STEINER, R. F.; *Trans. Faraday Soc.* 48 (1952) 1185
(116) STEFFENSEN, D.; *Proc. Nat. Acad. Sci. Wash.* 39 (1953) 613
(117) DARLINGTON, C. D.; *Proc. Roy. Soc. Lond.* B 118 (1935) 33–96; B 121 (1936) 264–319
(118) KUWADA, V. and NAKAMURA, T.; *Mem. Coll. Sci., Kyato Imp. Univ.* 9 (1933) 129

VOL 6 195

REFERENCES

(119) NAVASHIM, M.; *Univ. Calif. Publ. Agr. Sci.* 6 (1930) 95
(120) MANTON, I.; *Biol. Rev.* 25 (1940) 486
(121) MANTON, I. and SMILES, J.; *Ann. Bot. Lond., N.S.* 7 (1943) 195
(122) LILLIE, R. S.; *Biol. Bull.* 8 (1905) 193
(123) DARLINGTON, C. D.; *Recent Advances in Cytology*, 1939
(124) DARLINGTON, C. D.; *J. Genet.* 31 (1935) 185

VOL.
6
956

VOL
6
195

3

THE DONNAN EQUILIBRIUM

J. Th. G. Overbeek

CONTENTS

	Page
I. INTRODUCTION	58
II. ELEMENTARY THEORY	58
III. EQUILIBRIUM AND NON-EQUILIBRIUM PROPERTIES	61
IV. INTRODUCTION OF ACTIVITY COEFFICIENTS	63
V. APPROACH BASED UPON DETAILED CONSIDERATION OF THE UNEQUAL DISTRIBUTION OF IONS IN THE IONIC ATMOSPHERES	64
5.1 Theory	64
5.2 Qualitative treatment for the distribution of ions	64
5.3 Qualitative treatment of osmotic pressure	66
5.4 Quantitative treatment of the distribution of ions	67
5.5 Higher concentrations of the non-diffusible ions	69
5.6 Strong interaction of double layers	71
5.7 Activity coefficients	71
5.8 Experimental data	71
VI. THE DONNAN POTENTIAL	74
6.1 On the location of the Donnan potential	74
6.2 Formulation of the Donnan e.m.f.	75
6.2.1 Ideal solutions	77
6.2.2 Deviations from ideal behaviour	78
6.3 Experimental evidence	79
6.4 On the limited use of determining Donnan e.m.f., suspension effect or *p*H of a suspension	81
VII. SOME REMARKS ON BIOLOGICAL APPLICATIONS	82
REFERENCES	82

3

THE DONNAN EQUILIBRIUM

J. Th. G. Overbeek

1. Introduction

When two coexistent phases are subject to the restriction that one or several of the ionic components cannot pass from one phase into the other, a particular equilibrium is set up, called Donnan equilibrium. As a rule the restriction is caused by a membrane, permeable to the solvent and to small ions, but impermeable to ions of colloidal size, and therefore these equilibria are often called Donnan *membrane equilibria*. The presence of a membrane is, however, not essential. In a gel or a resin in which the structural component is charged, and even in a sediment of a substance like clay dissociating into very large and small ions the equilibria are of the Donnan type.

The Donnan equilibrium has three important aspects, viz.

1. The unequal distribution of ions.
2. The osmotic pressure.
3. The potential difference between the phases.

The elementary theory of the Donnan equilibrium[1] for ideal solutes is known well enough. As, however, electrolytes of colloidal size are usually far from ideal, it is important to investigate how this lack of ideality affects measurable properties. In the present article we shall therefore lay particular stress on the quantitative treatment of non-ideal systems.

2. Elementary Theory

The conditions for the Donnan equilibrium are conveniently described by means of the electrochemical potentials η_j introduced by Guggenheim[2]

$$\eta_j = \mu_j + z_j F \psi \qquad \ldots.(1)$$

where μ_j is the chemical potential of the species j, z_j its valency (with sign included), F the Faraday and ψ the electrical potential of the phase to which η_j and μ_j relate.

In equilibrium the temperature and the electrochemical potentials of all the diffusible components are equal in the two phases, but the pressures may, and as a rule do, differ:

$$\eta_j^{\,i} = \eta_j^{\,o} \qquad \ldots.(2)$$

where the two phases are distinguished by the superscripts i and o (inside and outside the membrane).

For ideal solutions these equalities can be written

$$\mu_j^s + p^i v_j + RT \ln x_j^i + z_j F \psi^i = \mu_j^s + p^o v_j + RT \ln x_j^o + z_j F \psi^o \quad \ldots.(3)$$

where it is assumed that the partial molar volume v_j of all components is independent of the pressure p. μ_j^s is the chemical potential in the standard state and x_j the mole fraction of species j.

The equations (3) together with the conditions for electro-neutrality (4) determine the state of the system completely.

$$\Sigma z_j x_j^i = 0 = \Sigma z_j x_j^o \quad \ldots.(4)$$

Specializing now for a completely dissociated uni-univalent salt, one non-diffusible (e.g. colloidal) ion of valency z and the solvent w, we obtain five equations

$$\begin{aligned}
(p^i - p^o)v_+ + F(\psi^i - \psi^o) &= RT \ln \frac{x_+^o}{x_+^i} \quad (5a)\\
(p^i - p^o)v_- - F(\psi^i - \psi^o) &= RT \ln \frac{x_-^o}{x_-^i} \quad (5b)\\
(p^i - p^o)v_w &= RT \ln \frac{x_w^o}{x_w^i} \quad (5c)\\
x_+^i - x_-^i + z x_z^i = 0 \quad x_+^o = x_-^o &= x \quad (5d, e)
\end{aligned} \quad \ldots.(5)$$

where the cation is indicated by $+$, the anion by $-$, the colloid ion by z and water by w.

The *distribution of ions* is found by adding eqs. (5a) and (5b)

$$\ln \frac{x_+^o x_-^o}{x_+^i x_-^i} = \frac{(p^i - p^o)(v_+ + v_-)}{RT} \quad \ldots.(6)$$

When the concentration of the colloid ion is not too large, the pressure difference $p^i - p^o$ is so small that the right-hand side of eq. (6) may be neglected.

The distribution of ions can then be calculated from eqs. (5d, e) and (7)

$$x_+^i x_-^i = x_+^o x_-^o \quad \ldots.(7)$$

or expressing in the same approximation the result in molar concentrations c instead of in mole fractions

$$c_+^i \cdot c_-^i = (c^o)^2 \quad \ldots.(8)$$

$$c_+^i - c_-^i + z c_z^i = 0 \quad \ldots.(9)$$

Here c^o is the salt concentration outside the membrane. The *osmotic pressure* follows easily from eq. (5c) in the following way

$$p^i - p^o = \frac{RT}{v_w} \ln \frac{x_w{}^o}{x_w{}^i} = \frac{RT}{v_w} \ln \frac{1 - 2x}{1 - x_+{}^i - x_-{}^i - x_z{}^i}$$

$$\approx \frac{RT}{v_w} (x_+{}^i + x_-{}^i - 2x + x_z{}^i)$$

$$\approx RT(c_+{}^i + c_-{}^i - 2c^o + c_z{}^i) \quad \ldots.(10)$$

which is simply Van't Hoff's expression.

The *potential difference* between the phases follows from eq. (5a) or (5b) after neglecting the small pv terms

$$\psi^i - \psi^o = \frac{RT}{F} \ln \frac{c_+{}^o}{c_+{}^i} = \frac{RT}{F} \ln \frac{c_-{}^i}{c_-{}^o} \quad \ldots.(11)$$

For comparison with approaches to be introduced later in this article, we shall give explicit equations for the extreme cases in which the colloid concentration is very much smaller or very much larger than the salt concentration viz. for $zc_z \ll c^o$ and for $zc_z \gg c^o$. (We drop the superscript i for the colloidal ion.)

Small colloid concentration

The solution of eqs. (8) and (9) including first order terms in $\frac{zc_z}{c^o}$ is

$$c_+{}^i = c^o - \frac{zc_z}{2} \quad \ldots.(12)$$

$$c_-{}^i = c^o + \frac{zc_z}{2} \quad \ldots.(13)$$

or, in words, *half of the charge of the non-diffusible ion is compensated by an excess of counterions and the other half by a shortage of the ions of the same sign as the colloid,* which we shall call the *co-ions.*

The osmotic pressure is equal to

$$p^i - p^o = RTc_z \quad \ldots.(14)$$

and the Donnan potential to

$$\psi^i - \psi^o = \frac{RT}{2F} \cdot \frac{zc_z}{c^o} \quad \ldots.(15)$$

The osmotic pressure has the same value as would be caused by the colloid ions alone.

Large colloid concentration

When $zc_z \gg c^o$, assuming that the colloid ion is positively charged we have

$$c_-{}^i = zc_z + \frac{(c^o)^2}{zc_z} \qquad \ldots.(16)$$

$$c_+{}^i = \frac{(c^o)^2}{zc_z} \qquad \ldots.(17)$$

The charge of the colloid ion is now nearly completely compensated by an excess of counterions.

The osmotic pressure is

$$p^i - p^o = RT\,(z + 1)c_z \qquad \ldots.(18)$$

and thus not only the non-diffusible ions but also the counterions are osmotically active.

The Donnan potential is given by

$$\psi_i - \psi_o = \frac{RT}{F} \ln \frac{zc_z}{c^o} \qquad \ldots.(19)$$

3. Equilibrium and Non-equilibrium Properties

A clear distinction should be maintained between the distribution of small ions and the osmotic pressure on one hand and the Donnan potential on the other hand. The first two are thermodynamically well-defined equilibrium properties.

The Donnan potential, however, can only be measured by introducing salt bridges into the two solutions and this adds irreversible processes to the equilibrium system. If, instead of using salt bridges, one would insert two identical reversible electrodes into the two solutions, zero potential difference would be found because the whole system would then be in equilibrium.

This fact, which has already been signalled by Donnan and Allmand,[3] is the basis of the so-called "indirect method" for determining the Donnan potential. Loeb[4], [5] has remarked that the Donnan potential is equal to the difference in pH between the two phases multiplied by 58 mV and often preferred the two pH measurements on the separate liquids to the insertion of salt bridges into the equilibrium system. The principle of this method is illustrated in Fig. 1. If the four electrodes (two hydrogen electrodes, and two salt bridges with calomel electrodes) are inserted into the Donnan system there is no potential difference between the two reversible electrodes, because no work can be obtained from a system in equilibrium. Taking the common potential of these two electrodes as zero, the Donnan potential

is evidently equal to $E_1 - E_2$ and on the other hand E_1 and E_2 are the e.m.f.'s from which the respective pH's are calculated. Consequently

$$p\mathrm{H}_{\text{susp.}} - p\mathrm{H}_{\text{equil. sol}} = \frac{E_{\text{susp.}} - E_{\text{equil. sol}}}{RT/0{\cdot}4343F} \qquad \ldots.(20)$$

In soil and suspension chemistry this difference in pH (or in the logarithms of the activities of any other ion) between a suspension and its intermicellar liquid is known as the suspension effect (WIEGNER, PALLMANN[6], [7], [8]). Recently its nature has formed the object of a

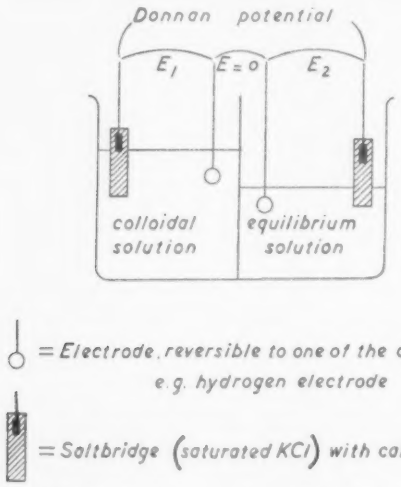


Fig. 1. Direct and indirect measurement of the Donnan potential. Demonstration of equality of Donnan potential and suspension effect.

rather heated discussion (JENNY *et al.*, MARSHALL, PEECH *et al.*, ERIKSSON, MYSELS, BABCOCK and OVERSTREET, SOLLNER, OVERBEEK, LOW[9–20]).

It will be clear that any conclusion on the Donnan potential can be carried over immediately to the suspension effect and vice versa.

The Donnan potential has been defined as the potential difference between the two *solutions*. The assumption that the Donnan potential is equal to the potential difference between the two salt bridges, implies that the liquid junction potentials are zero or at least of equal magnitude. As no independent means of determining potential differences between phases of different composition exist, this assumption can never be checked. Even if this is a reasonable approach for dilute solutions of simple electrolytes, its application to suspensions containing highly charged colloidal particles is open to doubt.

When considering non-ideal Donnan systems, the distribution of ions and the osmotic pressure, although more difficult to calculate, still

remain equilibrium properties. In the case of the Donnan potential, however, a detailed consideration of the two irreversible liquid junctions will prove essential.

4. Introduction of Activity Coefficients

In most cases Donnan equilibria do not obey the ideal laws. This is what we should expect considering that the non-diffusible ion is, as a rule, a highly charged colloidal particle which necessitates the introduction of large corrections for the deviations from ideal behaviour. These corrections can be given in the form of activity coefficients as has been done in a very rigorous way by Donnan and Guggenheim.[21]

Activity coefficients are introduced into equilibrium conditions of the type of eq. (3) by changing all x_j into $f_j \cdot x_j$ leading to:

$$\mu_j^s + p^i v_j + RT \ln f_j^i x_j^i + z_j F \psi^i = \mu_j^s + p^o v_j + RT \ln f_j^o x_j^o + z_j F \psi^o \quad \ldots.(21)$$

The condition of electroneutrality is unchanged

$$\Sigma z_j x_j^i = 0 = \Sigma z_j x_j^o \quad \ldots.(4)$$

Using the same treatment as in section 2, these conditions lead to the following equations.

For the distribution of ions we find instead of eq. (8):

$$f_+^i f_-^i c_+^i c_-^i = f_+^o f_-^o (c^o)^2 \quad \ldots.(22a)$$

or

$$(f_\pm^i)^2 c_+^i c_-^i = (f_\pm^o)^2 (c^o)^2 \quad \ldots.(22b)$$

The osmotic pressure is given by eq. (23), which replaces eq. (10):

$$p^i - p^o = \frac{RT}{v_w} \ln \frac{f_w^o x_w^o}{f_w^i x_w^i} = gRT(c_+^i + c_-^i - 2c^o + c_z^i) \quad \ldots.(23)$$

Here g is the conventional osmotic coefficient. Instead of eq. (11) the Donnan potential is given by:

$$\psi^i - \psi^o = \frac{RT}{F} \ln \frac{f_+^o c_+^o}{f_+^i c_+^i} \quad \ldots.(24)$$

Although the above equations are correct, they are of little practical value unless theoretical estimates for the activity coefficients can be given. An obvious approach for dilute systems is the use of Debye-Hückel activity coefficients. This has given satisfactory results in the case of the Donnan equilibrium KCl — $K_4Fe(CN)_6$ with a membrane impermeable to the ferrocyanide ion. This system has been investigated by Donnan and Allmand[3] and by Kameyama,[22] who

interpreted it with the help of LEWIS activity coefficients. HÜCKEL[23] showed that the results could also be explained with the help of Debye-Hückel activity coefficients. Quite recently, HILL[24] has shown that results equivalent to the use of Debye-Hückel coefficients are obtained by the application of the McMillan-Mayer theory of solutions, at least in the first approximation.

These theoretical approaches may be useful in the case of a relatively small non-diffusable ion or in the case of a swollen gel, where the fixed charges are distributed more or less uniformly. There are, however, serious drawbacks in the case of true polyelectrolytes where the non-diffusing charges are concentrated in a number of small spaces so that even with very great dilution of the polyelectrolyte molecules, the local concentration of ions remains high, close to the particle, and renders application of Debye-Hückel or similar theories impossible.

Instead of basing theoretical estimates of the activity coefficients on small departures of homogeneity as in the theory of DEBYE and HÜCKEL, a more promising approach is obtained by taking explicitly into account that the charges are concentrated in the colloidal particles. Such an approach has been given by DAVIS,[25] and more completely, by KLAARENBEEK[26] and will be treated in the next section. A short account of KLAARENBEEK's work has been given by the present author in KRUYT's *Colloid Science*.[27]

VOL
6
195

5. APPROACH BASED UPON DETAILED CONSIDERATION OF THE UNEQUAL DISTRIBUTION OF IONS IN THE IONIC ATMOSPHERES

5.1 *Theory*

As mentioned in section 4, the application of the Debye-Hückel approximation to the Donnan equilibrium often fails, because, for the high concentrations of ions close to the colloidal particles the substitution of $(1 - e\psi/kT)$ for $\exp(-e\psi/kT)$, as used in the Debye-Hückel theory, is not valid*.

In the Gouy-Chapman theory of the electrical double layer, as applied in the theoretical treatment of the electrocapillary curve (GRAHAME[28]), and in the theory of the stability of colloids (DERJAGUIN and LANDAU,[29] VERWEY and OVERBEEK[30]), this approximation is not used but the full exponential is preserved. It is worth while to do the same for the Donnan equilibrium.

5.2. *Qualitative treatment for the distribution of ions*

In a Donnan equilibrium the counterions are concentrated in the inner solution, and the co-ions are driven away from it. In accordance with

* Here ε is the ionic charge, ψ the local electrical potential, k BOLTZMANN's constant and T the absolute temperature.

eq. (21) this is understood as a consequence of the equality of electrochemical potentials in the two phases. The difference in electrical potential ($\psi^i - \psi^o$) between the two phases entails a corresponding difference in the concentrations of all the diffusible ions. In eq. (21) it is implicitly assumed that the electrical potential may be considered as uniform in each of the two phases.

It is now proposed to take the variations of the electrical potential in different parts of the suspension into account. This amounts to applying eq. (21) locally to each volume element, instead of to the average of the whole inner or outer phase.

In other words, in the classical treatment the *average* concentration of ions is assumed to be related to the *average* electrical potential, whereas we are now relating the *local* concentrations to the *local* potentials, the averages being taken later.

The local concentrations, e.g. in the immediate neighbourhood of a colloidal particle, are given by the BOLTZMANN theorem.

$$c_+ = c^o \exp(-F\psi/RT); \quad c_- = c^o \exp(+F\psi/RT) \quad \ldots.(25)$$

where c_+ and c_- are the local concentrations of the cations and the anions respectively and ψ is the local potential. c^o is the concentration of ions in the outer solution. The BOLTZMANN expression (25) can be regarded as the solution of eq. (21) or rather of eq. (3) when the local activity coefficients are all equal and the small pv terms are neglected.

The average concentrations of ions are:

$$\overline{c_\pm} = c^o \overline{\exp(\mp F\psi/RT)} \quad \ldots.(26)$$

and this is only equal to the concentration calculated with the average potential

$$\overline{c_\pm}' = c^o \exp(\mp F\bar{\psi}/RT) \quad \ldots.(27)$$

when the electrical potential is everywhere small enough to permit the replacement of $\exp(x)$ by $(1 + x)$.

If ψ is very small

$$\begin{aligned}\overline{c_\pm} &= c^o \overline{\exp(\mp F\psi/RT)} = c^o \overline{(1 \mp F\psi/RT)} \\ &= c^o (1 \mp F\bar{\psi}/RT) = c^o \exp(\mp F\bar{\psi}/RT) \quad \ldots.(28)\end{aligned}$$

If, however, the potential is high in the neighbourhood of the colloidal particles—and this occurs when the particles are highly charged and the ionic strength is low—then the accumulation of counterions is larger and the expulsion of co-ions smaller than corresponds to eq. (28). As the high potential close to the particles is nearly independent of the particle concentration, this dissymmetry persists *even in the case of very*

small colloid concentration, whereas according to the classical treatment the excess of counterions and the deficit of co-ions are the more nearly equal the lower the colloid concentration (cf. section 2). Fig. 2 illustrates the two cases.

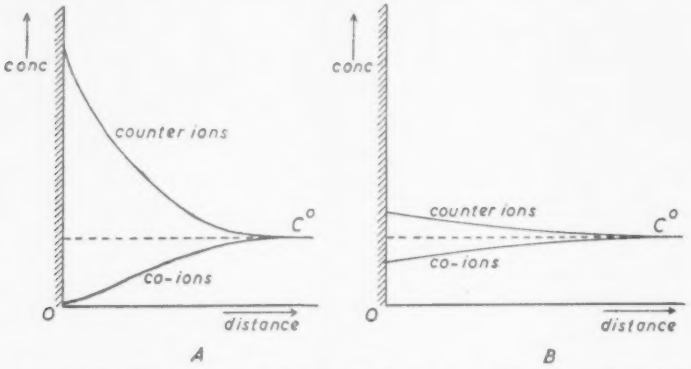


Fig. 2. Distribution of ions in the field of an electrically charged particle. *A*: for high potentials ($F\psi/RT > 1$). *B*: for low potentials ($F\psi/RT \ll 1$).

5.3. *Qualitative treatment of osmotic pressure*

In the classical treatment the osmotic pressure is determined by $c_+ + c_- - 2c^o + c_z$. When the inhomogeneity in the ionic concentrations is considered as in the foregoing subsection, one should consider that the pressure varies with the distance from the colloidal particles, the pressure being highest near the particles and lowest in certain

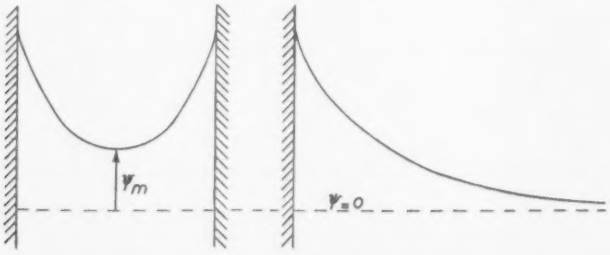


Fig. 3. Schematic representation of the course of the potential between two particles.

symmetry points midway between the particles. Eq. (3) or (5c) but now applied locally, allow the pressure to be calculated. The high pressures are partially counteracted by the stresses in the electric field and only in the symmetry point where the field strength is zero, may the pressure be considered as an osmotic pressure. In these symmetry points the potential still has a finite value (ψ_m) as shown schematically in Fig. 3. The pressure can then simply be calculated by

Van't Hoff's expression, which can be transformed with the aid of BOLTZMANN's theorem.

$$p^m - p^o = RT(c_+{}^m + c_-{}^m - 2c^o)$$
$$= RTc^o \{\exp(F\psi_m/RT) + \exp(-F\psi_m/RT) - 2\} \quad \ldots.(29)$$

where the index m stands for the symmetry points.

The osmotic pressure (rather, the contribution of the diffusible ions to the osmotic pressure) should be found by averaging expression (29) over all possible configurations. If the particles are not highly charged, so that mutual repulsion may be neglected, this averaging should lead to the classical value for the contribution of the small ions to the osmotic pressure

$$\bar{p} - p^o = RTc^o\{\exp(F\bar{\psi}/RT) + \exp(-F\bar{\psi}/RT) - 2\} \quad \ldots.(30)$$

If, however, the charges are high, the particles repel each other and approach a more lattice-like arrangement where the mutual distances are more or less identical and all p_m's are about equal, eq. (29) immediately giving the ionic contribution to the osmotic pressure.

As ψ_m in this regular arrangement is evidently smaller than the average value of the potential, the actual osmotic pressure is lower than the one calculated from average ionic concentrations. Experimentally this fact is known as the HAMMARSTEN effect.[31], [32]

5.4. *Quantitative treatment of the distribution of ions*

In order to give an accurate quantitative treatment based on the above ideas it would be necessary to take account of form, size and geometrical arrangement of the colloidal particles. As this is evidently beyond our actual powers of calculation, some approximation in this respect is necessary, but it is essential to conserve the possibility of using high values of the electrical potential. The most extensive treatment of the ionic atmospheres in the case of high potentials has been given for large flat faces. We follow therefore KLAARENBEEK's[26], [27] idea of representing the particles in a Donnan equilibrium as bounded by flat faces, neglecting any end effects.

In the first place we shall treat solutions, which are very dilute with respect to the colloidal particles, so that overlapping of double layers can be neglected. In this case the relative differences in the concentrations of ions in the two phases are very small. Instead of describing the distribution of ions by the ratio of the concentrations in the two phases, it can be more conveniently expressed as the fraction of the charge of the colloid that is compensated by an excess of counterions or by a deficit of co-ions respectively. This excess and deficit can be found by integrating the local concentrations over the available volume

and subtracting the bulk concentration taken over the same volume. The total charge is found as the sum of deficit and excess.

Consequently the fraction of the charge compensated by an excess of counterions, α_+, is equal to

$$\alpha_+ = \frac{\int_0^\infty [\exp(-F\psi/RT) - 1]\mathrm{d}x}{\int_0^\infty [\exp(-F\psi/RT) - \exp(F\psi/RT)]\mathrm{d}x} \quad \ldots.(31)$$

and similarly the fraction compensated by a deficit of co-ions, α_-, is given by

$$\alpha_- = \frac{\int_0^\infty [1 - \exp(F\psi/RT)]\mathrm{d}x}{\int_0^\infty [\exp(-F\psi/RT) - \exp(F\psi/RT)]\mathrm{d}x} \quad \ldots.(32)$$

where x represents the distance from the plane face of the particle. The particles are considered to be negatively charged and consequently ψ is always negative.

The integrations are straightforward and relatively simple when it is assumed that the electrical double layer is of the GOUY-CHAPMAN type[33], [34] (cf. VERWEY and OVERBEEK[30]).

The distribution of charges in the double layer obeys the Poisson-Boltzmann equation

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} = \frac{4\pi Fc}{\varepsilon}[\exp(F\psi/RT) - \exp(-F\psi/RT)] \quad \ldots.(33)$$

where c is the electrolyte concentration far from the particles and ε the dielectric constant of the solution.

A first integration leads to

$$\frac{\mathrm{d}\psi}{\mathrm{d}x} = -\sqrt{\frac{8\pi cRT}{\varepsilon}}[\exp(F\psi/2RT) - \exp(-F\psi/2RT)] \quad \ldots.(34)$$

where the fact that $\mathrm{d}\psi/\mathrm{d}x = 0$ at infinity has been taken into account. The proportionality between the surface charge density σ, and the field strength as given in eq. (35)

$$\sigma = -\frac{\varepsilon}{4\pi}\left(\frac{\mathrm{d}\psi}{\mathrm{d}x}\right)_{\text{surface}} \quad \ldots.(35)$$

leads to a simple relation between surface charge σ, and surface potential ψ.

$$\sigma = \sqrt{\frac{\varepsilon cRT}{2\pi}}[\exp(F\psi_0/2RT) - \exp(-F\psi_0/2RT)] \quad \ldots.(36)$$

where ψ_0 is the potential at the surface of the particle, i.e. at $x = 0$ while the potential is assumed to be zero at infinity.

The integral in eq. (32) for the co-ions can be transformed as follows:

$$\int_0^\infty [1 - \exp(F\psi/RT)]\,dx = \int_{\psi_0}^0 \frac{[1 - \exp(F\psi/RT)]}{d\psi/dx}\,d\psi \quad \ldots.(37)$$

Substituting $d\psi/dx$ from eq. (34) we find

$$\int_{\psi_0}^{0} \frac{[1 - \exp(F\psi/RT)]}{d\psi/dx}\, d\psi = \sqrt{\frac{\varepsilon}{8\pi cRT}} \int_0^{\psi_0} \frac{1 - \exp(F\psi/RT)}{\exp(F\psi/2RT) - \exp(-F\psi/2RT)}\, d\psi$$

$$= -\sqrt{\frac{\varepsilon}{8\pi cRT}} \int_0^{\psi_0} \exp(F\psi/2RT)d\psi = \sqrt{\frac{\varepsilon RT}{2\pi cF^2}}[1 - \exp(F\psi_0/2RT)] \quad \ldots.(38)$$

The integral in the denominator is handled similarly, leading to

$$\int_0^{\infty} [\exp(-F\psi/RT) - \exp(F\psi/RT)]dx$$

$$= \sqrt{\frac{\varepsilon RT}{2\pi cF^2}}[\exp(-F\psi_0/2RT) - \exp(F\psi_0/2RT)] \quad \ldots.(39)$$

The fraction of the particle charge compensated by a deficit of co-ions is thus

$$\alpha_- = \frac{\exp(F\psi_0/2RT) - 1}{\exp(F\psi_0/2RT) - \exp(-F\psi_0/2RT)} \quad \ldots.(40)$$

For very small values of the surface potential this fraction is equal to $\frac{1}{2}$ in agreement with Fig. 2*B*, but for larger values of ψ_0, i.e. for more highly charged particles, it decreases as shown in Table 1.

TABLE 1

Fraction of co-ions expelled (α_-) as related to the surface potential ψ_0

$\frac{F\psi_0}{RT}$	ψ_0* *in millivolts*	α_-
0	0	0·50
1	25	0·38
2	50	0·27
4	100	0·12
8	200	0·018

* At room temperature, $\frac{RT}{F}$ is close to 25 millivolt.

According to the elementary theory treated in par. 2 the fraction always goes to $\frac{1}{2}$, when the colloid concentration is lowered sufficiently (cf. eqs. 12, 13). In the present treatment the distribution remains asymmetrical, even at very high dilution.

5.5 *Higher concentrations of the non-diffusible ions*

Considering now higher particle concentrations where the double layers overlap significantly, we have to treat the case of two flat parallel faces carrying double layers as pictured in Fig. 3.

The overlapping of the double layers obviously makes the absolute value of the potential higher and cuts off the low potential tail of the double layer. Consequently the average values of $\exp(\mp F\psi/RT)$ become more extreme and the fraction of the particle charge compensated by the counterions will be larger, that compensated by the co-ions smaller. Qualitatively, the change is in the same direction as in the classical treatment; quantitatively it is less pronounced, because considerable interaction of double layers with high surface potentials is necessary before significant changes in the distribution of ions occur.

For a quantitative treatment we start again from the Poisson-Boltzmann equation (33). A first integration now leads to

$$\frac{d\psi}{dx} = -\sqrt{\frac{8\pi cRT}{\varepsilon}}\sqrt{2\cosh F\psi/RT - 2\cosh F\psi_m/RT} \qquad \ldots.(41)$$

where the boundary condition $\frac{d\psi}{dx} = 0$ half-way between the plates has been used.

The total counter charge $(-\sigma)$ per cm² of the double layer in one-half of the space between the two particles is given by

$$-\sigma = \frac{\varepsilon}{4\pi}\left(\frac{d\psi}{dx}\right)_{\text{surface}} = \sqrt{\frac{\varepsilon cRT}{2\pi}}\sqrt{2\cosh F\psi_0/RT - 2\cosh F\psi_m/RT} \qquad \ldots.(42)$$

The part of the total counter charge formed by the expulsion of co-ions can be written

$$\alpha_- = -\frac{Fc}{\sigma}\int_0^d (1 - \exp F\psi/RT)dx \qquad \ldots.(43)$$

After substituting $\frac{d\psi}{d\psi/dx}$ for dx this integral can be transformed into:

$$\alpha_- = \frac{1}{\sigma}\sqrt{\frac{\varepsilon cRT}{8\pi}}\int_{F\psi_0/RT}^{F\psi_m/RT}\frac{e^y - 1}{\sqrt{2\cosh y - 2\cosh(F\psi_m/RT)}}\,dy \qquad \ldots.(44)$$

The solution of the integral can be given with the aid of elliptic integrals, leading to

$$\alpha_- = \frac{[\{E(e^a, \pi/2) - E(e^a, \text{arc sin } e^{\frac{z-a}{2}})\} - (1 - e^a)\{F(e^a, \pi/2) - F(e^a \text{ arc sin } e^{\frac{z-a}{2}})\}]}{e^{a/2}\sqrt{e^z + e^{-z} - e^a - e^{-a}}} \qquad \ldots.(45)$$

where $a = F\psi_m/RT$, $z = F\psi_0/RT$, $F(k, \phi)$ is an elliptic integral of the first kind and $E(k, \phi)$ is an elliptic integral of the second kind in the notation of JAHNKE and EMDE.[35]

An application of these relations to the Donnan equilibrium demands the knowledge of ψ_m and ψ_o. These potentials are related through the charge density as expressed in eq. (42). If we assume the charge density to be a given quantity, one further relation is necessary. In principle this relation can be derived from the concentration of the colloid. This determines the distance $2d$ between the plates, and by a second integration of eq. (41) a relation between ψ_o, ψ_m and $2d$ can be established. However, KLAARENBEEK[26] felt that this might stretch the rather rough model too far and he therefore derived ψ_m from the experimentally determined osmotic pressure.

Eq. (46) shows how ψ_m is related to the osmotic pressure (cf. eq. 92).

$$p - p^o = RTc^o[\exp(F\psi_m/RT) + \exp(-F\psi_m/RT) - 2] \ldots . (46)$$

With ψ_m and σ known, ψ_o can be calculated from eq. (42) and all the data necessary to evaluate α_- are given.

5.6. *Strong interaction of double layers*

If the particles come very close together the potential drop between ψ_o and ψ_m becomes smaller and smaller as illustrated in Fig. 4. This means that the local concentrations of ions are not very variable and that we are again back at the classical Donnan situation, where distribution of ions, osmotic pressure and Donnan potential are simply

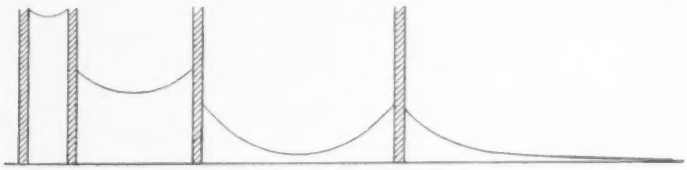


Fig. 4. Illustrating schematically that the range of potentials between two particles becomes relatively smaller, the closer the particles are together.

given by the elementary equations as treated in section 2. It is remarkable that the new treatment of the Donnan effects may deviate strongly from the classical treatment when the colloid concentration is low, but not when it is high.

5.7. *Activity coefficients*

It would be possible to calculate activity coefficients corresponding to the treatment given in this section, but such a calculation would have no great advantages. Activity coefficients are introduced in order to give account of deviations of classical laws. They are strictly empirical quantities. As soon as a theoretical explanation is put forward the introduction of activity coefficients does not throw any fresh light on the problem.

5.8. *Experimental data*

The theoretical picture outlined in the preceding sections, has been drawn in order to explain a number of deviations from ideal behaviour. Of the existing experimental data, those of KLAARENBEEK[26] on the Donnan equilibrium of gum arabic with KBr are best suited to test the theory and to show its advantages over the classical approach.

KLAARENBEEK used solutions of Na arabinate[26] with a molecular weight of about 200·000 and an equivalent weight of 1180. Potassium bromide was used as the diffusible salt because the Br^- ion can be

determined relatively easily. KLAARENBEEK determined the distribution of ions by measuring the Br^- content inside and outside the membrane, the osmotic pressure and in a number of cases the Donnan potential. His results on the distribution of ions can be most easily expressed as the amount of Br^- expelled: cf. eq. (40).

The drawn curves in Fig. 5 are experimentally determined lines. The dotted lines are valid for ideal behaviour or for equal mean activity coefficients in the two phases. It is seen that, especially at the lower concentrations of KBr, the deviations are striking. They do not disappear with diminishing concentrations of gum arabic. This fact, that even at very high dilution of the colloidal particles the Donnan theory for ideal behaviour is not followed, is one of the strongest indications of the necessity of considering a dilute solution of a colloid as a number of islands of high concentration of ions dispersed in a solution of a much lower concentration, rather than as a completely homogeneous system.

The experimental slopes at zero gum arabic concentration in 0·001 0·01, 0·1 and 1 *N* KBr solutions are 0·11; 0·19; 0·32 and 0·50 g eq. Br^- expelled per g eq. gum respectively.

Using eq. (40), these slopes allow the calculation of the surface potential and with the aid of eq. (36), of the apparent surface charge density.

TABLE 2

Surface Potential and Surface Charge Density of the Gum Arabic Particles as calculated from the Distribution of Ions

Conc. of KBr *in g eq. per litre*	*g eq.* Br^- *expelled per g eq. gum arabic*	*Surface potential* ψ_0 *in mV*	*Surface charge density* σ *in e.s.u. per cm²*
0·00108	0·115	102	− 4300
0·01	0·19	72	− 7100
0·1	0·32	37	− 8700
1·0	0·50	(0)	(0)

The order of magnitude of potential and charge density is not unreasonable; neither is the fact that the potential decreases and that the charge density increases with decreasing concentration of salt (the value 0 for σ at the highest salt concentration is extremely sensitive to error in the analytical determination).

Considering for a moment the gum arabic molecule (M.W. 220·000, Equiv. W. 1180) as a spherical particle with 185 elementary charges, a charge density of 7000 e.s.u./cm² would correspond to a surface area of $1{\cdot}3 \times 10^{-11}$ or a radius of 10^{-6} cm, which is quite a reasonable figure,

corresponding well with a value of $1{\cdot}1 \times 10^{-6}$ cm from diffusion (BRUINS[37]) and of about 10^{-6} cm from titration curves (OVERBEEK[38]).

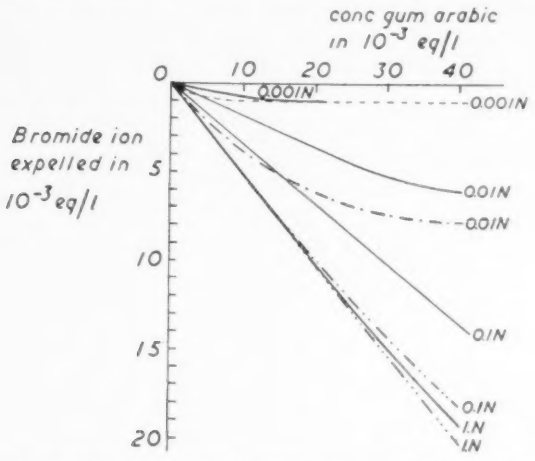


Fig. 5. Expulsion of Br^- ions in the Donnan equilibrium between gum arabic and KBr for four different concentrations of the outside solution. The experimental curves are drawn. The dotted curves represent ideal behaviour.

The curves for 1 N and 0·1 N KBr are practically straight lines, indicating the absence of measurable overlapping of double layers.

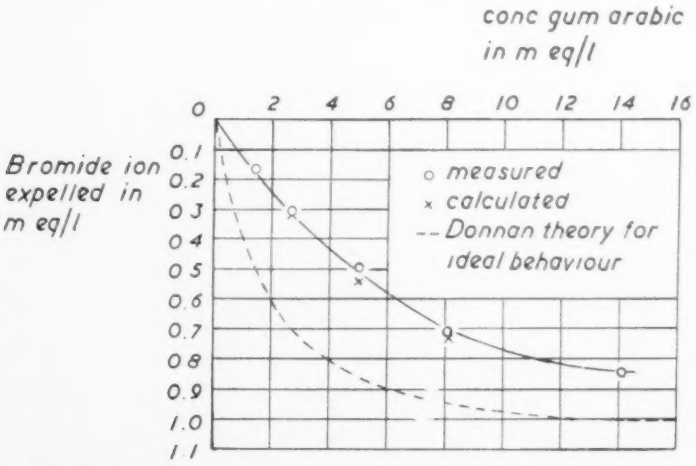


Fig. 6. Expulsion of Br^- ions from solutions of gum arabic with 0·00108 N KBr, as measured (0), calculated from eq. (45) (×), and calculated for ideal behaviour (dotted line).

The curvature of the 0·01 N and 0·001 N lines indicates that at higher concentrations of gum arabic some interaction of the double layers can

be detected, resulting in a still smaller expulsion of co-ions. Quantitatively this should be explained by the theory as indicated in section 5.5. It has been mentioned there already that given the difference between the model (flat faces) and the actual situation (coiled molecules), a straightforward calculation of the effect of higher concentrations of the colloid is hardly possible. KLAARENBEEK has therefore calculated the value of the potential half-way between the plates, from the osmotic pressure with the aid of eq. (46), and assuming a surface potential independent of the concentration of gum he has calculated the amount of co-ions expelled. Numerical data are given in Table 3 and shown for 0·001 *N* KBr in Fig. 6. The good agreement between calculated and experimental values is another indication that although the model may be rough the line of thought indicated in this section merits serious consideration.

TABLE 3

Experimental and Calculated Values for the Expulsion of Co-ions at Higher Concentrations of Gum Arabic

Conc. KBr	*Eq. conc. gum arabic*	*Osmotic pressure in cm* H_2O	*Potential half way in mV calculated with eq.* (46)	*Surface potential in mV* (from Table 2)	Br^- *ion expelled in g eq. per g eq. gum arabic*	
					Calculated from eq. (45)	*Found*
0·00108 N	$1{\cdot}4 \times 10^{-3}$ N	0·3	2·8	102	0·119	0·118
,,	2·7 ,,	1·9	6·8	,,	0·116	0·113
,,	2·0 ,,	8·7	14·6	,,	0·108	0·099
,,	8·1 ,,	32·9	27·7	,,	0·090	0·087
,,	14·1 ,,	12·7	39·2	,,	0·071	0·060
0·01 N	$2{\cdot}7^5 \times 10^{-3}$ N	1·35	1·8	72	0·197	0·17
,,	5·5 ,,	2·1	2·2	,,	0·195	0·193
,,	11·0 ,,	6·5	4·0	,,	0·194	0·192
,,	20·3 ,,	23·8	7·8	,,	0·189	0·181
,,	30 ,,	74·1	13·6	,,	0·176	0·170

6. THE DONNAN POTENTIAL

6.1. *On the location of the Donnan potential*

It has already been mentioned in section 3, that the Donnan potential and the e.m.f. connected with the suspension effect are identical, and that saturated KCl bridges play an important role in their measurement. In the usual interpretation of this measurement the liquid junctions with saturated KCl are assumed to be potential free, or nearly so, the measured e.m.f. being located nearly exclusively at, or in the membrane between the two Donnan phases. It is, however, by no means certain

that the diffusion potential between saturated KCl and a colloidal solution or a suspension is small. In a liquid junction between saturated KCl and a dilute solution of electrolytes of low molecular weight, the contribution to the diffusion potential of the concentration gradient of the dilute electrolyte is negligible, because it is drowned in a large excess of KCl. The concentration gradient of the KCl does not contribute to the diffusion potential because the mobilities of K^+ and Cl^- are nearly the same. In the liquid junction between saturated KCl and a colloidal solution or suspension, the diffusion gradient of the colloid still plays a negligible part on account of the high excess of KCl. In the tail of the concentration gradient of KCl, however, where the equivalent concentration of KCl is of the same order as that of the colloidal ions, the mobilities of K^+ and Cl^- may be considerably changed in the high electric fields around the particles. The changes for K^+ and Cl^- are in opposite directions and thus may lead to a considerable diffusion potential. It has even been argued by JENNY *et al.*[9] and by BABCOCK and OVERSTREET,[17] that the Donnan potential is completely situated at this liquid junction and not at the membrane. However, we run up here against the well-known difficulty of defining the potential difference between two phases of different composition. GUGGENHEIM[39], [40] showed that this quantity is not accessible to any experimental determination.

VOL. 6 1956

There is a way out of this difficulty, as a theory can be formulated on the e.m.f. of the complete Donnan cell (including the two liquid junctions) in which these equivocal potentials of separate phases are avoided (OVERBEEK[19]).

It has been proposed that this e.m.f. which is accessible to measurement, be called the Donnan e.m.f., as distinct from the Donnan potential between the two phases, which cannot be measured.

6.2. *Formulation of the Donnan e.m.f.*

The essential part of the cell with the aid of which the Donnan e.m.f. is usually determined, consists of the two Donnan phases in equilibrium, each in contact with a saturated solution of KCl. It is irrelevant to which type of reversible

$$\underset{\text{sat. KCl}}{\text{I}} \;\Big|_{1}\; \underset{\text{equilibrium solution}}{\text{II}} \;\Big|_{2}\; \underset{\text{suspension}}{\text{III}} \;\Big|_{3}\; \underset{\text{sat. KCl}}{\text{IV}} \qquad (47)$$

membrane

electrodes the saturated KCl phases are connected. The membrane is permeable to water and to all the small ions and supposedly, completely impermeable to the suspended particles.

The total e.m.f. of the cell[47] can be calculated as the sum of three

liquid junction potentials. Applying the usual equation for the *l.j.* potential (cf. MACINNES[41]) one finds

$$-EF = \int_{\text{left}}^{\text{right}} \sum_i \frac{t_i}{z_i} \, \mathrm{d}\mu_i \qquad \ldots.(48)$$

where the sum is to be taken over all the ions, including the colloidal particles. t_i is the transference number, z_i the valence with sign included, and μ_i the chemical potential of the species i. In order to transform the chemical potentials of single ions occurring in (48), into chemical potentials of neutral systems, we add to (48) the obvious equality (49).

$$0 = \int_{\text{left}}^{\text{right}} \sum_i t_i \mathrm{d}\mu_K \qquad \ldots.(49)$$

where K represents a monovalent positive reference ion, say the potassium ion.

The integral in (49) is zero because $\Sigma t_i = 1$ and μ_K has the same value in the two saturated KCl solutions.

In this way we find

$$-EF = \int_{\text{left}}^{\text{right}} \sum_i t_i \left(\frac{\mathrm{d}\mu_i}{z_i} - \mathrm{d}\mu_K \right) \qquad \ldots.(50)$$

The expression in parentheses refers to a neutral substance. If i represents a negative ion,

$$\frac{\mathrm{d}\mu_i}{z_i} - \mathrm{d}\mu_K = \frac{1}{z_i} (\mathrm{d}\mu_i - z_i \mathrm{d}\mu_K) = \frac{1}{z_i} \mathrm{d}\mu_{K_z i}$$

where $K_z i$ is the salt consisting of one i-ion and z K^+-ions. If i is a positive ion $\dfrac{\mathrm{d}\mu_i}{z_i} - \mathrm{d}\mu_K$ represents the change in free energy accompanying an ion exchange of one K-ion by $1/z_i$ i-ion.

Eq. (50) can be further transformed by splitting the integral into three integrals relating to the three phase boundaries in the cell (47).

$$-EF = \int_{\text{sat. KCl}}^{\text{equil. sol}} \sum_i t_i \left(\frac{\mathrm{d}\mu_i}{z_i} - \mathrm{d}\mu_K \right) + \int_{\text{equil. sol}}^{\text{suspension}} \sum_i t_i \left(\frac{\mathrm{d}\mu_i}{z_i} - \mathrm{d}\mu_K \right) + \int_{\text{suspension}}^{\text{sat. KCl}} \sum_i t_i \left(\frac{\mathrm{d}\mu_i}{z_i} - \mathrm{d}\mu_K \right) \qquad \ldots.(51)$$

As the two Donnan phases are in equilibrium, the chemical potentials of all diffusible salts are equal in these two phases. The chemical potential of a salt of which the non-diffusible ions form a component, changes across the membrane but the transference number of the non-diffusible ions is obviously zero in the membrane. Consequently all the

terms in the sum in the second integral of eq. (51) are zero and the equation for the Donnan e.m.f. reduces to

$$-EF = \int_{\text{sat. KCl}}^{\text{equil. sol}} \sum_i t_i \left(\frac{\mathrm{d}\mu_i}{z_i} - \mathrm{d}\mu_K\right) + \int_{\text{suspension}}^{\text{sat. KCl}} \sum_i t_i \left(\frac{\mathrm{d}\mu_i}{z_i} - \mathrm{d}\mu_K\right) \quad \ldots.(52)$$

This equation which has a clear thermodynamic significance, not only shows that the e.m.f. of the Donnan cell can be completely expressed in the properties of the solutions at the two liquid junctions, but also that this e.m.f. is connected with transference numbers as well as with chemical potentials.

In order to evaluate the Donnan e.m.f. in practical cases, the relations between transference numbers and chemical potentials in the liquid junctions have to be ascertained. Although strictly, the course of the concentrations of all the components in the liquid junctions has to be known, a very reasonable approximation is obtained by assuming that in the significant part of the *l.j.* the colloidal particles and all the other constituents of the Donnan system have a constant concentration, the only variation being a gradient of the concentration of potassium chloride.

6.2.1. *Ideal solutions*

If all the solutes, including the non-diffusible ions, behave ideally both with respect to mobility and to activity, the following relations hold.

$$\left.\begin{aligned} \sum_{i \neq K,Cl} t_i &= 1 - t_K - t_{Cl} \\ \mathrm{d}\mu_i = RT\frac{\mathrm{d}c_i}{c_i};\ \mathrm{d}\mu_K &= RT\frac{\mathrm{d}c_K}{c_K} \\ \mathrm{d}c_K &= \mathrm{d}c_{Cl} \end{aligned}\right\} \quad \ldots.(53)$$

Substituting the relations (53) into eq. (52) leads to:

$$\begin{aligned} -EF &= \int_{\text{sat. KCl}}^{\text{equil. sol}} (1 - t_K - t_{Cl})(-\mathrm{d}\mu_K) + t_{Cl}(-\mathrm{d}\mu_{Cl} - \mathrm{d}\mu_K) \\ &\quad + \text{similar expression for the second integral} \\ = -EF &= RT \int_{\text{sat. KCl}}^{\text{equil. sol}} \left(-\frac{\mathrm{d}c_K}{c_K} + \frac{t_K \mathrm{d}c_K}{c_K} - \frac{t_{Cl}\mathrm{d}c_{Cl}}{c_{Cl}}\right) \\ &\quad + RT \int_{\text{suspension}}^{\text{sat. KCl}} \left(-\frac{\mathrm{d}c_K}{c_K} + \frac{t_K \mathrm{d}c_K}{c_K} - \frac{t_{Cl}\mathrm{d}c_{Cl}}{c_{Cl}}\right) \\ = -EF &= RT \int_{\text{sat. KCl}}^{\text{equil. sol}} \left[\left(\frac{t_K}{c_K} - \frac{t_{Cl}}{c_{Cl}}\right)\mathrm{d}c_K - \mathrm{d}\ln c_K\right] \\ &\quad + RT \int_{\text{suspension}}^{\text{sat. KCl}} \left[\left(\frac{t_K}{c_K} - \frac{t_{Cl}}{c_{Cl}}\right)\mathrm{d}c_K - \mathrm{d}\ln c_K\right] \end{aligned}$$

As the mobilities of K^+ and Cl^- are very nearly equal, the transference numbers of these ions are in nearly the same ratio as their concentration. Thus the terms $(t_K/c_K - t_{Cl}/c_{Cl})$ are very close to zero and the Donnan e.m.f. reduces to

$$E = \frac{RT}{F}\left[\int_{\text{sat. KCl}}^{\text{equil. sol}} \mathrm{d}\ln c_K + \int_{\text{suspension}}^{\text{sat. KCl}} \mathrm{d}\ln c_K\right] = \frac{RT}{F}\ln\frac{c_{K\ \text{equil. sol.}}}{c_{K\ \text{suspension}}} \quad \ldots.(54)$$

which is identical to eq. (11). This shows that the treatment *via* the liquid junction potentials is equivalent to the classical treatment, provided the approximations for ideal behaviour are used.

6.2.2. *Deviations from ideal behaviour*

In order to apply eq. (52) to non-ideal cases, both transference numbers and chemical potentials have to be known as functions of the composition of the solutions in the liquid junctions. In the present state of our knowledge of these quantities, the best we can do is to use a very simple example, i.e. a solution of a potassium salt of a polyelectrolyte ion P with valence z and potassium chloride, and make reasonable estimates of mobilities and activities.

Specializing eq. (52) to the case just mentioned one obtains:

$$E_{\text{Donnan}} = \frac{1}{F}\int_{\text{sat. KCl}}^{\text{equil. sol}} t_{Cl}\mathrm{d}\mu_{KCl} + \frac{1}{F}\int_{\text{suspension}}^{\text{sat. KCl}}\left(t_{Cl}\mathrm{d}\mu_{KCl} + \frac{t_P}{z}\,\mathrm{d}\mu_{PK_z}\right) \quad \ldots.(55)$$

The term $\frac{t_P}{z}\,\mathrm{d}\mu_{PK_z}$ is usually small. We will neglect its influence here, but if wanted it can be expressed in terms of $\mathrm{d}\mu_{KCl}$ and the composition of the system as has been shown by OVERBEEK.[19] In the two integrals μ_{KCl} has the same limits. Eq. (55) can thus be written

$$E_{\text{Donnan}} = \frac{1}{F}\int_{\mu_o}^{\text{sat. KCl}} (t_{\text{Cl(susp.)}} - t_{\text{Cl(equil. sol.)}})\mathrm{d}\mu_{KCl} \quad \ldots.(56)$$

in which μ_o is the chemical potential of KCl in suspension and equilibrium solution.

Assuming the mobilities of K^+ and Cl^- to be equal the transference number in the equilibrium solution is $\frac{1}{2}$. In the suspension the transference number differs from $\frac{1}{2}$ for two reasons. In the first place there are more K^+ ions than Cl^- ions (we neglect here the contribution to the conduction by the non-diffusible ions). In the second place the mobilities of K^+ and Cl^- in the neighbourhood of the colloidal particles are modified. Especially the mobility of the counterions may be far below their normal value as is proved by the low equivalent conductivity

of completely dialyzed suspensions or polyelectrolyte solutions. A reduction of the mobility by a factor four is not exceptional.[19], [42] The mobility of the co-ions will not be greatly changed because they are driven away from the regions of high charge density.

If the non-diffusible ions carry a high charge (and this is the case when deviations from ideal behaviour will be most obvious), it will not be a bad approximation to assume that this charge is completely compensated by an excess of counterions, the concentration of co-ions remaining the same as in the equilibrium solution (cf. section 5.4). Assuming moreover that the mobility of the co-ions is completely unaffected and that the mobility of the counterions compensating the charge of the non-diffusible particles is reduced to $1/r$ of its normal value, the transference number of the Cl ion in the suspension is given by

$$t_{Cl} = \frac{c}{2c + zc_z \frac{1}{r}} \qquad \ldots.(57)$$

where c is the concentration of KCl in equilibrium with the suspension and c_z is the molar concentration of particles.

Writing $2RT \, \mathrm{d} \ln c$ for $\mathrm{d}\mu_{KCl}$, eq. (56) is now easily integrated leading to

$$E_{\text{Donnan}} = \frac{RT}{F} \ln \frac{c}{c + \frac{zc_z}{2r}} \qquad \ldots.(58)$$

For a dilute suspension this is approximately equal to

$$E_{\text{Donnan}} = -\frac{RT}{2F} \cdot \frac{zc_z}{c} \cdot \frac{1}{r} \qquad \ldots.(59)$$

This expression, being identical to the ideal expression (15) except for the factor $1/r$ shows that the first and biggest effect on the Donnan e.m.f. is given by the non-ideal behaviour of the mobilities rather than of the activities.

6.3. *Experimental evidence*

A complete test of the theoretical considerations given in this section cannot yet be given because the necessary transference data are lacking. We can only take some existing data on Donnan-potentials and compare interpretations with the theory just described and with the classical Donnan approach.

KLAARENBEEK[26] determined Donnan e.m.f.'s in systems containing sodium arabinate and KBr. The e.m.f. was proportional to the concentration of gum, with proportionality constants of 3·5, 0·7 and

0·07* mV per milli equivalent of gum for KBr equilibrium concentrations of 0·001, 0·01 and 0·1* N respectively. In a 1 N solution there was no measurable e.m.f. Interpretation of these data with the classical equation

$$E = \frac{RT}{2F} \frac{zc_z}{c^o} \qquad \ldots.(15)$$

would lead to the conclusion that the degree of dissociation of Na (or K) arabinate is equal to 30 per cent, 60 per cent and 60 per cent* in 0·001, 0·01 and 0·1 N KBr respectively. It does not seem reasonable to suppose that the dissociation should *increase* with increasing content of salt. The interpretation with eq. (52) or approximately with eq. (59) would mean that in the lower electrolyte concentration the mobility of the cations is considerably decreased and this is wholly in line with the fact that at low ionic strengths the electrical potential around the particles is high.

Similar conclusions can be drawn from data on the Donnan e.m.f. of sodium bentonite published by DAVIS[25] in which the apparent degree of dissociation increases from 2·5 per cent to 10 per cent when the NaCl concentration increases from 0·001 to 0·01 N.

ADAIR and ADAIR[43], [44], [45] made very careful and highly accurate measurements of the Donnan e.m.f. of a number of protein systems and of Congo red. In all cases they found a linear relationship between concentration of colloid and e.m.f. provided that the latter was not more than a few millivolts. They considered this as suggesting that the activity coefficients of the diffusible salts are the same in the two phases and therefore as an indication of the applicability of eq. (15). We have seen, however, that the linearity by itself is a consequence of either theory. It is therefore desirable to compare the valence of the colloid as calculated with eq. (15) with independent evidence. Measurements on Congo red,[43] a disodium salt of a disulphonic acid which is aggregated in solution, interpreted with eq. (15) seem to show that only 50 per cent of the sodium ions are dissociated. Unfortunately the data are given for one ionic strength (0·02) only, so that the course of this degree of dissociation with concentration of salt is not known. However, the assumption that the mobility of the sodium counterions is decreased by 50 per cent seems preferable to the assumption of a very incomplete dissociation of a sodium salt of a sulphonic acid in dilute solution (0·02 M).

ADAIR's data on protein, viz. edestin, serum albumin, hemoglobin[43] and egg albumin[45] and the recent data by CHARLWOOD[46] on horse serum albumin all show a relatively good agreement of the valence of these proteins calculated from eq. (15) with independent estimates of the valence, e.g. from analytical data, from osmotic pressures and from

* The figures for the 0·1 N solution are rather uncertain because E is so small.

electrophoretic mobilities.[46], [47] As most of these experiments have been performed at relatively high ionic strengths, the electrical potentials in the ionic atmospheres were rather low, which explains the good agreement. Larger deviations from the classical theory are expected especially for low salt contents.

A remarkable effect that has been observed in studies on the suspension effect in clays[14], [48] is a reversal of the sign of the suspension effect without a corresponding change in sign of the electrophoretic mobility of the suspended particles. The effect seems to be connected with the presence of bi- or polyvalent counterions. An explanation based on the classical theory seems very difficult. In eq. (52), however, in which as we have seen the counterions usually play a predominant part, the transference number of the counterions needs only to become sufficiently negative to explain a reversed suspension effect or Donnan potential. Negative transference numbers in colloidal systems are not abnormal. Cf. data by HARTLEY, COLLIE and SAMIS[49] on cetyl pyridinium bromide.

VOL. 6 1956

6.4. *On the limited use of determining Donnan e.m.f., suspension effect or p*H *of a suspension*

We have seen that the Donnan e.m.f. and the suspension effect are determined by a rather complicated mixture of mobilities and activities of all the ionic species in the system. Simple relations only exist when the whole system behaves ideally or somewhat less stringently when the electrical potential near the non-diffusible particles remains low ($\psi \ll RT/F$). Only in this case can the Donnan potential be interpreted in a simple way, giving information about the valency of the non-diffusible particles (cf. eq. 15). If the condition of low potential is not fulfilled, however, interpretation is difficult even in the limit of vanishing concentration of colloid. In these circumstances more information can be derived from separate determinations of activities or mobilities.

If this point of view is accepted, one should also realize that the same kind of difficulty may be encountered in the interpretation of the *p*H of a suspension or polyelectrolyte solution. Much of the value of a *p*H measurement derives from the assumption that *p*H gives a reasonable estimate of the activity of the hydrogen ions. This can never be exact because one deals here with a single ion activity but as long as there are reasons to assume that the liquid junction potential with saturated KCl is zero or at least independent of the hydrogen ion activity, the calculation of a_H from *p*H makes some sense. With polyelectrolytes present in relatively large amounts any such assumption on the *l.j.* potential becomes of doubtful value and an interpretation of the *p*H in terms of activities of the H^+ ions loses its sense.

It may be good to quote here words used by P. B. TAYLOR[50] in 1927

in a treatise on *p*H measurement: ". . . for the cell e.m.f. depends not only on the acid activity but also on the activity of every molecular species in the cell and mobility of every ion. If these are sufficiently well known to be allowed for, the acid activity is likely to be sufficiently well known not to need measurement."

7. Some Remarks on Biological Applications

This paper was aimed at a treatment of some difficulties in non-ideal Donnan equilibria rather than at giving specific applications to biological systems. Three remarks may, however, well be made.

(*a*) *Qualitatively* there is little difference between the classical approach and that presented here. The suspension has the higher osmotic pressure, the counterions are accumulated in it and quite often to a higher degree than the co-ions are driven away. The sign of the Donnan e.m.f. is the same as the sign of the charge of the particles. Only from the last rule exceptions may be found and explained as suggested in section 6.3. Only if results are to be interpreted *quantitatively*, should a critical attitude prevail. Perhaps unexpectedly the largest deviations from classical Donnan theory may be expected with the most dilute equilibrium solutions.

(*b*) Donnan systems are (or should be) in equilibrium. In many biological systems membranes separate phases, that are *not* in equilibrium, so that Donnan theory does not apply.

(*c*) When electrical potentials are measured in biological systems, quite frequently liquid junctions are used to make contact with different parts of the system, e.g. the inside and outside of a nerve fibre. For obvious reasons saturated potassium chloride cannot be used in the liquid junction. A more dilute solution containing far less K-ion is used, e.g. a suitable Ringer solution. The treatment as given in section 6 can be applied to such systems with hardly any modification. It should be realized that the remarkable effect produced by the saturated KCl bridges in the Donnan cell are nearly completely determined by the low concentration side of the liquid junctions and consequently will also be present in these potentials measured in biological systems.

References

(1) Donnan, F. G. and Harris, A. B.; *J. Chem. Soc.* 99 (1911) 1554
Donnan, F. G.; *Z. Elektrochem.* 17 (1911) 572; *Chem. Revs.* 1 (1924) 73

(2) Guggenheim, E. A.; *J. Phys. Chem.* 33 (1929) 842

(3) Donnan, F. G. and Allmand, A. J.; *J. Chem. Soc.* 105 (1914) 1941

(4) Loeb, J.; *Proteins and the Theory of Colloidal Behaviour*, McGraw-Hill, 2nd ed., p. 178 ff (1924)

(5) Loeb, J.; *J. Gen. Physiol.* 3 (1921) 85, 547, 667, 691

(6) Wiegner, G.; *Kolloid-Z.* 51 (1930) 49

(7) Wiegner, G. and Pallmann, H.; *Z. Pflanzenernähr. Düng. u. Bodenk.* A 16 (1930) 1

VOL.
6
1956

REFERENCES

(8) PALLMANN, H.; *Kolloidchem. Beihefte* 30 (1930) 334

(9) JENNY, H., NIELSEN, T. R., COLEMAN, N. T. and WILLIAMS, D. E.; *Science* 112 (1950) 164

(10) COLEMAN, N. T., WILLIAMS, D. E., NIELSEN, T. R. and JENNY, H.; *Soil Sci. Soc. Amer. Proc.* 15 (1951) 106, 111, 114

(11) MARSHALL, C. E.; *Science* 113 (1951) 43; 115 (1952) 361; *Soil Sci. Soc. Amer. Proc.* 15 (1950) 110; 17 (1953) 218, 221

(12) PEECH, M. and MCDEVIT, W. F.; *Soil Sci. Soc. Amer. Proc.* 15 (1950) 112

(13) PEECH, M. and SCOTT, A. D.; *ibid.* 15 (1950) 115

(14) PEECH, M., OLSEN, R. A. and BOLT, G. H.; *ibid.* 17 (1953) 214, 220

(15) ERIKSSON, E.; *Science* 113 (1951) 418

(16) MYSELS, K. J.; *Science* 114 (1951) 424; 118 (1953) 603

(17) BABCOCK, K. L. and OVERSTREET, R.; *Science* 117 (1953) 686

(18) SOLLNER, K.; *J. Colloid Sci.* 8 (1953) 179

(19) OVERBEEK, J. TH. G.; *J. Colloid Sci.* 8 (1953) 593

(20) LOW, P. F.; *Soil Science* 77 (1954) 29

(21) DONNAN, F. G. and GUGGENHEIM, E. A.; *Z. Physik. Chem.* A 162 (1932) 346

(22) KAMEYAMA, N.; *Phil. Mag.* (6), 50 (1925) 849

(23) HÜCKEL, E.; *Kolloid-Z.* 36 (Zsigmondy Festschrift) (1925) 204

(24) HILL, T. L.; *J. Chem. Phys.* 22 (1954) 7, 1251; cf. also HILL, T. L. in CLARKE, H. T. (Ed.) *Iontransport Across Membranes.* Academic Press, New York, 1954, p. 210 ff

(25) DAVIS, L. E.; *Soil Science* 54 (1942) 199; 59 (1945) 379

(26) KLAARENBEEK, F. W.; Over Donnan evenwichten bij solen van Arabische Gom, Thesis, Utrecht, 1946

(27) OVERBEEK, J. TH. G.; in KRUYT, H. R. (Ed.) *Colloid Science*, Vol. I, Elsevier, Amsterdam, 1952, pp. 191–193

(28) GRAHAME, D. C.; *Chem. Revs.* 41 (1947) 441

(29) DERJAGUIN, B. and LANDAU, L.; *Acta Physicochim. USSR* 14 (1941) 633; *J. Expt. Theor. Physics* (Russ.) 11 (1941) 802; 15 (1945) 662

(30) VERWEY, E. J. W. and OVERBEEK, J. TH. G.; *Theory of the Stability of Lyophobic Colloids*, Elsevier, Amsterdam, 1948

(31) HAMMARSTEN, E. and H.; *Arkiv. Kemi. Min. Geol.* 8 (1923) 1
HAMMARSTEN, H.; *Biochem. Z.* 147 (1924) 481
HAMMARSTEN, E.; *Biochem. Z.* 144 (1924) 383

(32) RIJSSELBERGHE, P. VAN; *J. Phys. Chem.* 38 (1934) 645

(33) GOUY, G.; *J. Physique* (4) 9 (1910) 457; *Ann. Phys.* (9) 7 (1917) 129

(34) CHAPMAN, D. L.; *Phil. Mag.* (6) 25 (1913) 475

(35) JAHNKE, E. and EMDE, F.; *Tables of Functions with Formulae and Curves*, 4th ed., Dover Publications, New York, 1945, p. 52 ff

(36) BUNGENBERG DE JONG, H. G. and THEUNISSEN, P. H.; *Kolloid. Beih.* 47 (1938) 245

(37) BRUINS, H. R.; *Kolloid.-Z.* 57 (1931) 152

(38) OVERBEEK, J. TH. G.; *Bull. Soc. Chim. Belges* 57 (1948) 252

(39) GUGGENHEIM, E. A.; *J. Phys. Chem.* 33 (1929) 842

(40) GUGGENHEIM, E. A.; *Thermodynamics*, North Holland Publ. Co., Amsterdam, 1950, p. 332

(41) MACINNES, D. A.; *The Principles of Electrochemistry*, Reinhold, New York, 1939, p. 221 ff

(42) OVERBEEK, J. TH. G.; in KRUYT, H. R. (Ed.), *Colloid Science*, Vol. I, Elsevier Amsterdam, 1952, p. 238 ff

(43) ADAIR, G. S. and ADAIR, M. E.; *Biochem. J.* 28 (1934) 199, 1230

(44) ADAIR, G. S. and ADAIR, M. E.; *Trans. Faraday Soc.* 31 (1935) 130

(45) ADAIR, G. S. and ADAIR, M. E.; *Trans. Faraday Soc.* 36 (1940) 23
(46) CHARLWOOD, P. A.; *Biochem. J.* 46 (1950) 312
(47) TISELIUS, A. and SVENSSON, H.; *Trans. Faraday Soc.* 36 (1940) 16
(48) RAUPACH, M.; *Austral. J. Agric. Res.* 5 (1954) 716
(49) HARTLEY, G. S., COLLIE, B. and SAMIS, C. S.; *Trans. Faraday Soc.* 32 (1936) 795
(50) TAYLOR, P. B.; *J. Phys. Chem.* 31 (1927) 1499

VOL
6
195

4

BIOLOGY AND BIOPHYSICAL PROPERTIES OF TRANSFORMING PRINCIPLES

Stephen Zamenhof

CONTENTS

		Page
I.	INTRODUCTION	86
II.	THE BIOLOGICAL PHENOMENON	87
	(*a*) Species Susceptible to Transformation	88
	(*b*) Features Transferable in the Transformation Phenomenon	90
	(*c*) The Nature of the Transforming Phenomenon	92
	(*d*) The Transduction Phenomenon	96
III.	THE TRANSFORMING PRINCIPLE	96
	(*a*) The Chemical Nature of the Transforming Principle	97
	(*b*) The Chemical Analysis	100
	(*c*) The Heterogeneity	101
	(*d*) The "Native State" of DNA	102
	(*e*) The Isolation and Purification of the Transforming Principles	103
	(*f*) Physical properties of the Transforming Principle	104
	(*g*) The Molecule of the Transforming Principle	105
	(*h*) The Resistance to Physical and Chemical Agents	107
	1. Heat	107
	2. H^+ and OH^- ions	108
	3. Deoxyribonuclease	109
	4. Ionic Strength	110
	5. Dehydration	110
	6. Deamination	111
	7. Mutagenic Agents	111
	8. The Unstable Transforming Principle	115
	REFERENCES	116

4

BIOLOGY AND BIOPHYSICAL PROPERTIES OF TRANSFORMING PRINCIPLES

Stephen Zamenhof

I. Introduction

This review deals with one biological phenomenon, transformation, and with one substance, the transforming principle. The phenomenon is comparatively little known, its precise nature not being agreed upon; thus far it has been observed only in bacteria, and only in a few species at that. What, then, makes the discovery of this phenomenon one of the milestones in biological sciences?

Towards the end of the first half of this century the foundations for chemical explanation of several biological phenomena had already been laid. However, the phenomenon most essential for life—heredity—was not attractive to a chemist, probably because of a fear of the multitude of substances involved and of their insurmountable complexity; indeed, in contrast to simpler biochemical functions, nothing less than the whole cell or at least the entire chromosomal apparatus seemed indispensable as heredity determinant. The remainders of this attitude still hamper modern research on the chemistry of the transmission of heredity.

The discovery of the transforming phenomenon and of the chemical nature of the transforming principles furnished the first possibility of a chemical attack. As will be discussed further in this review, the transforming principles seem indeed to be heredity determinants (although it is not known whether or not they are the only or the most important heredity determinants); they can be extracted from the cell, purified, chemically identified and analysed; they can be chemically changed *in vitro*, re-introduced into the cell and studied to establish the relationship between the change in structure and the change in function; in short, they permit the application to the problem of heredity of the same rational approach which yielded elucidations in so many fields of biochemistry and biophysics.

While the above mentioned reasons for interest in the transforming principle appear to the author to be the most important ones, the biophysicist's interest may often come from another direction. A considerable amount of work has been done in the field of a conspicuous cell constituent, deoxyribonucleic acid (DNA). In early chemical approach the DNA was subjected to degradation, with complete disregard of the macromolecular nature of the substance. This

error was compensated for after the last war by extensive biophysical studies of the giant molecule of DNA; but the biophysicists soon faced a dilemma: whether their substance was indeed "native" or was still degraded. The discovery that some purified highly polymerized DNA have transforming activity but lose it on the slightest degradation, provided the desired yardstick of "intactness"; for although no one can say whether his DNA preparation is in the same state as the DNA that exists in the cell, the "functionally intact" unit seems important enough to warrant study, and constant enough to serve as a standard.

The third reason for interest in the transforming principle is a practical one: subtle reactions of certain agents, notably mutagenic, carcinogenic and carcinostatic agents, with the DNA, can often be demonstrated by studying the loss of transforming activity, long before these reactions can be apprehended by any chemical or physical method. On the other hand, several carcinogenic and carcinostatic agents have no effect on the transforming activity which may serve as some indication that the primary action of these substances is through cell components other than DNA.

All the above points will be discussed in the following chapters. Chapter II will be concerned with the biological phenomena themselves, valid regardless of the nature of the transforming principle; Chapter III will contain discussion of the nature and properties of the transforming principle, valid regardless of the transforming phenomena which will serve here merely as detectors of activity.

II. The Biological Phenomenon*

In 1928 Griffith[4] reported that when mice were injected with living non-encapsulated (R) pneumococci mixed with heat-killed encapsulated (S) ones, living S pneumococci were recovered from the animals. Of particular interest was the fact that the specific type of the S cells was the same as the type of the heat-killed cells rather than the type from which the living R cells were derived (by some previous mutation to the loss of capsule). Once the new feature (production of the capsule of a specific type) was established, it was retained and reproduced in subsequent generations as if a new gene had been added to the genetical makeup of the receptor cells; indeed, such transformed cells, when heat-killed, could now serve to induce the transformation of R cells exactly in the same way as did the original S cells.

Although this experiment was confirmed by other workers, it did not attract due attention: despite careful controls it still appeared possible that the seemingly heat-killed preparation contained some heat-resistant bacterial forms so that the phenomenon was a "revival" of

* For previous reviews on this subject see [1], [2], [3].

the S cells rather than the "transformation" of the R cells. The possibility of a heat-resistant virus also was not excluded.

In 1931 DAWSON and SIA[5] reported that transformation can also be demonstrated *in vitro*. These workers realized that the phenomenon affects only one out of many R cells and that it is necessary to depress the growth of this strain to give the few newly formed S cells a chance to multiply to the point of recognition. They therefore performed the experiment in the presence of an anti-R serum; today this method is still the preferred one when transforming the R cells into S cells.

ALLOWAY[6] was first to demonstrate that the presence of the whole "donor" (S) cells is not necessary for the phenomenon: this investigator prepared cell-free aqueous extracts of the heat-killed S cells, passed the extract through a bacterial filter and demonstrated that the filtrate is still capable of transforming the R cells into S cells. While this experiment excluded the presence of a non-filterable bacterial form, it did not exclude the filterable virus as the responsible agent; the experiment was merely suggestive that the agent is water-soluble.

In 1944, AVERY, MACLEOD and MCCARTY[7] purified the extract further and found that the responsible agent (the "transforming principle") has all the properties of a highly polymerized deoxypentose nucleic acid (DNA). This single finding laid the foundation for the study of the transforming principle; it will be discussed in greater detail in the next chapter.

The post-war period witnessed several important contributions to our knowledge of the phenomenon and its ramifications; they will be discussed below under appropriate headings.

(*a*) *Species susceptible to transformation*

To date, reproducible transformation phenomena are known only in bacteria. A transformation-like phenomenon in viruses has been reported by BERRY and DEDRICK[8] and by BERRY.[9] These investigators have changed the virus of rabbit fibroma into that of infectious myxomatosis by means of heat-killed myxoma virus. This finding has been confirmed by others;[10], [11], [12], [13] however, it is difficult to judge whether the nature of this phenomenon is similar to that of bacterial transformations.

The transformation of organisms higher than bacteria has never been reported. The swift progress in the field of tissue culture makes it predictable that attempts to transform mammalian cells will soon be made.

In bacteria, the organism used originally by GRIFFITH, the pneumococcus (or more properly *streptococcus pneumoniae*) is still most widely used, because of the large body of knowledge accumulated, the comparatively low pathogeneicity of this organism, and, above all, the

reproducibility of the results. The disadvantage of this species is the presence in the bacterial cultures of the enzyme, deoxyribonuclease (DNase), which tends to destroy the transforming principle; this subject will be fully discussed in the next chapter.

In 1945, BOIVIN *et al.*[14], [15], [16] reported transformation in *Escherichia coli*. However, after BOIVIN's death, the phenomenon could not be repeated by his collaborators and by numerous experienced investigators in U.S.A. This is not surprising if one realizes that, even if the transforming phenomenon can be demonstrated in a species, only certain strains (or mutants) of this species are susceptible as "receptor" strains; thus, in a species of high variability*, such as *Escherichia coli* (compare[17]), a susceptable strain might easily and irreversibly be lost. This may also explain frequent failures to reproduce the results of other reported transformations. It may be significant that the "dependable" transformation phenomena have thus far been demonstrated only in the parasites which live in a rather non-changeable medium, such as mammalian organs; in such species, a high variablity would be a disadvantage for the species and the unstable genes probably have been eliminated by natural selection.[18]

Other bacterial species, in which transformation was reported but not yet confirmed, are in chronological order: *Shigella paradysenteriae*,[19] *Proteus* OX19,[20] Salmonella,[20], [21] Staphylococcus[23] and the tubercle bacillus.[22] More promising are results on transformation of *Alcaligenes radiobacter* and *Phytomonas tumefaciens*[23], [24] and of *Brucella*.[25]

Besides pneumococcus, the only two species in which the transformation phenomenon can be repeated day after day with the same reliability, are *Hemophilus influenzae* (ALEXANDER and LEIDY[26]) and *Neisseria meningitidis* (ALEXANDER and REDMAN[27]). *Hemophilus influenzae*† lends itself particularly well to quantitative studies on the transforming principle, and indeed most of the quantitative work has been done on this species. The reason for this is a practically complete absence of the DNase from bacterial cultures, as evidenced by the fact that the activity of the introduced transforming principle does not decrease in the culture.[29] Another advantage of this species over pneumococcus is that in *Hemophilus influenzae* no additional "sensitizing agents" are necessary for successful transformation. This point will be fully discussed in the next chapter.

The question why transformations have been demonstrated in only a few species is frequently asked. One reason may be the presence of

* The term "variability" is used here to denote the net result of interaction of two mechanisms: mutation and selection.

† This transformation was soon confirmed by an independent laboratory (TUNEVALL[28]).

powerful DNases in the cultures of many species, such as *Serratia marcescens*[30] or group A β-hemolytic streptococci; if not inhibited, these DNases may destroy the introduced transforming principle before it has had a chance to penetrate into the receptor cell. Another reason may be the non-susceptibility of the receptor strain used; this phenomenon, not well understood, will be discussed later on; it may involve, among other things, the non-permeability of the cell to the transforming principle. The third reason may be seen in the human factor: success in the search for a new transforming phenomenon requires thorough familiarity with the organism used, as well as with the properties of DNA (transforming principle) and DNases; years of patient work may also be required.

(*b*) *Features transferable in the transformation phenomenon*

The feature transferred (induced) in the transformation phenomenon as originally discovered by GRIFFITH[4] was the production of capsules type I, II or III of pneumococcus. (Each capsule contains a polysaccharide specific for its type.) Further studies of MCCARTY and AVERY[31] proved that the production of capsules of types VI or XIV (chemically different from I, II and III) can also be induced by using the transforming principle from the cells of these types. It is probable that transformation could be demonstrated for every one of the 70 and odd known types of pneumococcus. Because of the fact that early investigations of the transformation phenomenon were limited to the production of various polysaccharides, some investigators suggested that the transforming principle may be merely a "primer" of a polysaccharide nature. This objection became meaningless when it was found that a multitude of other features can also be transferred by transformation. AUSTRIAN and MACLEOD[32] demonstrated acquisition of a specific M protein through transformation in pneumococci. The six capsular substances *a*, *b*, *c*, *d*, *e* and *f*, produced in the transformation of *Hemophilus influenzae* (ALEXANDER and LEIDY[26]) appeared first to be polysaccharides but closer chemical study (ZAMENHOF *et al.*[33], [34]) revealed that most of them form a new class of immunologically active compounds, the poly-sugarphosphates. The substance of type *b* is of particular interest, as it appears to consist of a poly-ribophosphate chain such as exists in pentose nucleic acids, in which the place of the purines and pyrimidines is occupied by a second similar chain, linked to the first in 1:1′ glycosidic linkages. The substance of type *a* appears to be a poly-glucophosphate, that of type *c*—a poly-galactophosphate. The substance of type *d* does not contain phosphate[35] but is highly negatively charged.[36] The transferable feature in the transformation of *Neisseria meningitidis* (ALEXANDER and REDMAN[27]) is the production of capsular substance type I or II*a*; the chemical natures of these

VOL
6
195

substances have not been well investigated, but the substances appear to be more complex than polysaccharides.

Although the feature transferred is probably in most cases the capacity to produce one or more specific enzymes, in the right amounts, times and locations, many transferable features at present are not necessarily traceable to any definite enzymes or other substances. MACLEOD and KRAUSS[37] and TAYLOR[38] described transformation in *Pneumococci* in which the transferred features were the productions of different *quantities* of polysaccharide (type III), resulting in a different *morphology* of colonies. A suggestion has been made[38] that the genes involved form indeed a series analogous to what is known in higher organisms as *allelic* series. AUSTRIAN[39] reports the transformations to eighteen (out of the twenty predicted) morphologic variants in pneumococcus. In one recently reported case (EPHRUSSI-TAYLOR[40]), the change in appearance (large colonies) was traced to a change in the metabolism of glucose and lactic acid.

Other transferred features were the inductions of fermentation of salicin[41] or mannitol;[42] in such experiments again one cannot be sure *a priori* whether the transformation affected the production of the fermenting enzyme, the permeability of the cell or some other unknown factor; the fermentation of mannitol, however, has been traced down to the presence of an active mannitol phosphate dehydrogenase.[43], [44]

An important class of transferred features, still less traceable to definite substances, is the acquisition of resistance or sensitivity to antibiotics. HOTCHKISS[45] was first to demonstrate that penicillin-sensitive strains of pneumococcus can be transformed into penicillin-resistant. In this experiment the penicillin, rather than antiserum, is used as a selective agent to depress the growth of the non-transformed cells. The induction of penicillin-resistance was achieved stepwise: each successive treatment with the transforming principle from the fully-resistant donor resulted in an increase of penicillin-resistance of the transformed receptor cells. These results are interesting also because they closely resemble the stepwise acquisition of penicillin-resistance by a natural process that is the spontaneous mutations.[46] HOTCHKISS reported also transformation to resistance to streptomycin[45] and to sulfanilamide.[42] Full resistance to streptomycin can be achieved in one transformation step; this situation is again analogous to one of the natural processes of acquiring resistance to streptomycin by spontaneous mutation. Recently, HOTCHKISS and MARMUR[42] reported an interesting transformation from streptomycin-resistance to streptomycin-sensitivity.

The phenomenon of transformation to resistance to antibiotics is not limited to pneumococcus. ALEXANDER and LEIDY[47] demonstrated transformation to streptomycin-resistance in *Hemophilus influenzae*;

the one-step acquisition of full streptomycin-resistance was also demonstrated for this organism.

(c) *The nature of the transforming phenomenon*

Although, as discussed above, a considerable amount of work has been done in the field of bacterial transformations, the nature of the phenomenon itself remains far from clear. If one approaches the phenomenon with an unbiased mind and if one is considering solely the experimental evidence, the following conclusions appear logical: The molecules of the transforming principle are hereditary determinants because their presence determines the presence of hereditary characters. However, it is not known whether *all* hereditary characters can be accounted for by the (joint) action of all the molecules of transforming principle in the cell;[3] it is still conceivable that the determination of the most fundamental features proceeds through an entirely different mechanism.

The transforming principle seems to be molecules of DNA (this point will be fully discussed in the next chapter); on successful transformation they must reproduce (or be reproduced) because more of them are obtained. They could reproduce either outside (for instance on the surface) or inside the cell. If the first were true, the DNA having transforming activity would be found outside the intact cell. This, however, is contrary to the experimental evidence: as will be shown in the next chapter, a thorough destruction of the bacterial cell is necessary to isolate the DNA (whether or not it has demonstrable transforming activity). Furthermore, within three minutes after addition of transforming principle to the cells, the former becomes completely protected against the action of added strong DNase (ALEXANDER and LEIDY[26]); this phenomenon cannot be easily explained if the transforming principle remained on the outside of the cell. Thus it appears that, on transformation, the molecule of DNA must penetrate the cell. But the mechanism of this hypothetical penetration remains unknown: the DNA is composed of giant molecules and ought to be stopped by a normal bacterial membrane; at neutral *p*H it is highly negatively charged and ought to be repelled by the majority of bacterial cells which are also known to be negatively charged. Recent finding (ALEXANDER and LEIDY[48]) that the *encapsulated* strains *b* and *d* (strong negative charge[33], [36]) are susceptible to transformation in no lesser degree than the non-encapsulated ones, make the penetration still more mysterious, both for permeability and repulsion reasons. One is tempted to speculate that since only a small proportion of receptor cells is susceptible to transformation, these cells are in several respects different from the normal ones.

The problem of susceptibility to transformation is still intriguing the investigators in this field. As mentioned before, DAWSON and SIA[5]

VOL
6
195

found that serum containing antibodies against the original (R) strain was necessary for successful transformation of pneumococcus. These workers, however, thought this to be the only function of the serum. AVERY, MACLEOD and MCCARTY[7] were the first to realize that serum may be necessary also for reasons other than the presence of antibodies. Indeed, the anti-R serum is not necessary in transformation to resistance to antibiotics; yet even in these experiments, the presence of some additional factors was necessary for transformation of pneumococcus. Closer investigation[49], [38], [50] established that these additional factors in serum comprise (*a*) a dialysable factor replaceable by pyrophosphate ions; (*b*) a protein fraction replaceable by fraction V of bovine serum albumin. These factors are believed to produce "sensitization" of receptor cells, a temporary change of unknown nature; a change in bacterial surface to admit the transforming principle has been suspected here. Even when the receptor cells are "sensitized," in practice only about 1 per cent of them undergoes transformation,[3] and the addition of excess of transforming principle does not improve this figure. It has been, however, claimed (RAVIN[51]) that every sensitized cell is potentially capable of being transformed.

These conditions apply thus far to pneumococcus only. In the transformation of *E. coli*[14], [15], [16] and in the better studied transformations of *Hemophilus influenzae*[26] and of *Neisseria meningitidis*[27] no need for additional factors (other than those in the medium) could be established; these factors, therefore, do not seem essential for the transformation phenomenon. In *Hemophilus influenzae* the percentage of cells undergoing transformation is 10 to 10^3 times lower than in pneumococcus. Recent extensive study of the susceptibility to transformation in this organism (ALEXANDER, LEIDY and HAHN[52]) revealed that: (*a*) the percentage of cells which can become susceptible is an inherited characteristic of the strain used, but there is no evidence that the susceptible cells reproduce their kind; (*b*) the proportion of susceptible cells very rapidly increases towards the end of logarithmic growth, to reach a peak in the early stationary phase of the growth cycle, and declines thereafter. Essentially similar studies have been made by HOTCHKISS.[53] This author concludes that the groups of susceptible cells are successfully developing and disappearing as the culture grows.

In summary, then, the susceptibility to transformation appears to require a special, otherwise unknown, physiological state of the receptor cell. This susceptibility may mean not only the admission of transforming principle into the cell, but its further acceptability inside the cell. It is logical to assume that the transforming principle (DNA) upon admission does not remain loose (in solution). If analogy can be drawn to higher organisms, bacteria may also possess chromosomes

(DE LAMATER[54]) which contain all* the DNA of the cell, in the form of nucleoproteins, in the structures corresponding to genetical "loci": only chromosomal DNA may enjoy the benefit of orderly reproduction through chromosomal apparatus. Thus upon entering the susceptible cell, the DNA may become fixed to the "locus," probably by precipitating as nucleoprotein. That some sort of "fixation" or precipitation seems necessary has been postulated recently by demonstration that the molecule of DNA of *E. coli* is in solution actually longer than the cell of *E. coli*†. (ROWEN and NORMAN[55]).

Another phenomenon, which may involve both admission to the cell and the acceptability inside the cell, is the phenomenon of competition. ALEXANDER, LEIDY and HAHN[52] have demonstrated that when the transforming principle from *H. influenzae* cells type *a* was mixed, in a certain proportion, with one from type *b*, and the mixture used to transform R cells, the transformed cells represented a mixture of cells of type *a* and *b*, in a proportion similar to that used for preparation of the mixture of transforming principles *a* and *b*. Similar results were obtained when mixing the transforming principles *b* and *c*. Under appropriate experimental conditions, prior exposure to high concentration of transforming principle of one type can completely exclude subsequent transforming principle from effecting transformation. These findings can be taken as evidence that various transforming principles compete for the susceptible cell and, in this competition, seem to obey the law of mass action.

This competition can be visualized as competition for entrance into the cell and/or competition for the "locus" on the "chromosome," to mention only two of many possibilities. It stands to reason that the competition for the "locus" should be very specific whereas the competition for entrance into the cell (or "blocking of the cell surface") should be also demonstrable when the blocking agent is non-specific. Non-specific competition was demonstrated in *H. influenzae*[52] using calf thymus DNA as the blocking (competing) agent. However, this phenomenon is weaker than specific competition and the complete exclusion of transforming principles cannot be achieved even by a 1000-fold excess of calf thymus DNA. Similar blocking by calf thymus DNA was also observed by HOTCHKISS[53] in pneumococcus.

The problem of hypothetical competition for the "locus" may be closely connected with a problem of the DNA exchange.

At least three logical possibilities should be taken into consideration when attempting to explain the acceptance of the transforming principle in the susceptible receptor cell. For simplicity, only type-transformation will be discussed here. One possibility is that the

* With few exceptions (e.g. eggs).

† The cell of *Hemophilus influenzae* is still smaller.

VOL
6
195

accepting "locus" (protein?) was devoid of any DNA and therefore the gene was incomplete and non-functional; thus, the cell remains typeless (R). The acceptance of the transforming principle of a specific type completes the gene (nucleoprotein?) for this type, and an encapsulated (S) cell results. A second possibility is that the locus is never devoid of DNA and that it always forms a complete gene; but in the R type this gene produces none or very little of the capsular substance. On transformation, this DNA would be replaced by the DNA for production of demonstrable amounts of type-specific substances; this DNA may be introduced in excess, and the reaction involved may be an exchange reaction obeying the law of mass action. A third possibility is that the DNA becomes accepted in some other unknown way and that it makes itself felt by modifying the activities of the already existing genes. This list, of course, does not pretend to be complete; it merely helps to establish some working hypothesis. That the introduction of a transforming principle may abolish the action of an already existing gene has been shown by TAYLOR;[38] this phenomenon was taken as a demonstration of "allelism" of the genes involved. The finding that an S cell of one type can be transformed into an S cell of another type directly (i.e. without passing through R type) (ALEXANDER and LEIDY[48]) also points in this direction. All these experiments may seem to favour the second possibility (DNA exchange), but admittedly very little is still known about the mechanism of the transforming phenomenon; for that matter, it cannot even be said that all the transforming phenomena have the same mechanism.

That the mechanisms of some transformation phenomena may indeed be different from the classical ones is strongly suggested by the discovery of transformations in which the transformed cells are *unlike* the donor cells from which the transforming principle was derived. The discovery was made independently by EPHRUSSI-TAYLOR[56] in pneumococcus and by LEIDY, HAHN and ALEXANDER[57] in *Hemophilus influenzae*. In pneumococcus, the addition of transforming principle of type SIII-2 to the receptor cells of type SIII-1 resulted in production of not only type SIII-2 cells, but also of cells of type S_N which is distinctly different from SIII-1 and SIII-2. It was postulated that the transforming principles SIII-1 and SIII-2 interact on reproduction in a manner somewhat resembling recombination of bacteriophages or chromosomal crossing-over in higher organisms. These transformations in which the outcome is unlike the donor cells, have been called "allogenic," as distinct from the classical transformations called here "autogenic." Both types have been further studied by RAVIN.[51] In *Hemophilus influenzae*, the addition of transforming principle from type *a* to the receptor cells type *b* (or R cells divided from smooth type *b*), produces, among other cells, the cells S_{ab}, which never existed in

nature before: each of these cells produces capsular substances of both types *a* and *b*; the cells reproduce new transforming principle *ab* which is not just a mixture of transforming principles *a* and *b*. It appears likely that the new transforming principle *ab* is a result of a "faulty" reproduction of *a* in presence of *b* (or vice versa). This interesting phenomenon, which could be also called a "directed mutation," will be fully discussed further on.

(*d*) *The transduction phenomenon*

LEDERBERG, LEDERBERG, ZINDER and LIVELY[58] noticed that when Salmonella cells with certain hereditary traits are lysed, especially by weakly lytic phages, the lysate contains a filterable agent capable of inducing these traits in other Salmonella cells; the phenomenon has been given the name "transduction."

The cause of the phenomenon was eventually found to be associated with a bacteriophage[59], [60] presumably serving as a carrier for the material which changes the heredity of the host (receptor) cells. The nature of this phenomenon is still not clear. If the hereditary material transferred is a chemical substance (DNA) then the phenomenon can be called a special case of transformation (or vice versa); however, if, as appears probable (DEMEREC[61]) the transferred material is a chromosome or its part then the phenomenon is of an entirely different nature. Moreover, for this study it should be stressed that the hereditary material, when outside of the cell, seems to be completely protected against the action of DNase[58]; this indicates that the phenomenon does not lend itself to physico-chemical study of the material *in vitro*. For this reason, the phenomenon, though undoubtedly very interesting, will not be discussed further in this review.

III. THE TRANSFORMING PRINCIPLE

As outlined before, in this chapter the transforming phenomenon will be considered merely as a detector of activity of the transforming principle. However, the transforming phenomenon involves several steps, and at present it is not possible to decide which is (are) responsible for inactivation in any particular case: thus, if a transforming principle loses its ability to transform it might be due to any of the following (hypothetical) failures: the inability to penetrate into the cell, to reach the genetical locus, to displace the previous occupant (DNA) of the locus, to attach itself to the protein of the locus, to reproduce itself, (even though established in the locus), to exert its phenotypical action, even though reproduced, etc.

The chemistry and physics of the DNA itself have been the subject of many reviews (compare[62]; see also volumes III and IV of this

series[63], [64]). In this section, therefore, these subjects will be discussed only with reference to the transforming principle.

(*a*) *The chemical nature of the transforming principle*

As mentioned before, in 1944, AVERY, MACLEOD and MCCARTHY[7] found that the purified transforming principle has all the properties of a highly polymerized DNA. Their conclusion that the transforming principle *is* DNA was based on the following observations: (1) Elementary analysis of the transforming principle corresponds to that of DNA; (2) Chemical tests reveal the presence of DNA; no other substance could be detected by either chemical or physical (ultracentrifuge, electrophoresis) methods; (3) Serological tests fail to detect the presence of any immunologically active substances (such as polysaccharides); (4) Of several enzymes tested, only (crude) deoxyribonuclease (DNase) was able to destroy the transforming activity. At the same time, the physical study (viscosity, ultracentrifugation) revealed that the substance is in a highly polymerized state. These findings, especially destruction by DNase, were amplified by further studies of MCCARTY and AVERY.[65], [31]

The conclusion that the transforming principle is DNA was not immediately accepted: the workers in the field were still influenced by the authority of "old masters" (especially P. A. LEVENE[66]) who, on the basis of erroneous chemical analysis, concluded that the DNA from all sources contain equimolar amounts of individual purines and pyrimidines; that these form sub-units ("tetranucleotides"); and that, in fact, all DNA are identical, being composed of identical "tetranucleotides"; of course, the identical and simple DNA molecules could not serve as highly specific heredity determinants. The erroneous views about the structure of DNA were slowly corrected, mainly through the work of CHARGAFF and his collaborators (for a review, see[62]); however, some criticism of the identification of the transforming principle with DNA still remained valid.[67] The main points of criticism were: (1) the chemical methods were not sensitive enough to exclude the presence of an impurity (protein?) which might be responsible for the activity; (2) even accepting the evidence that the destruction of DNA by DNase destroys the activity, it could still be that DNA is not active alone, but merely in combination with such a hypothetical protein so that the destruction of either moiety results in inactivation; indeed, for a long time the genes were suspected of being *nucleoproteins*, in which various proteins would be carriers of specificity whereas undiversified DNA—merely an unspecific prosthetic group; (3) the non-destruction of activity by a few proteolytic enzymes by itself does not prove the non-protein nature of the transforming principle because proteins are known which resist many proteolytic enzymes.

To answer the first point of criticism, attempts were made in recent years to improve methods of purification and analysis. ZAMENHOF *et al.*[68], [69] purified the transforming principle of *H. influenzae* to the point that it contained less than 0·4 per cent (referred to DNA) of each of the following impurities: proteins, immunologically active substances, and ribonucleic acids. No loss of biological activity occurred

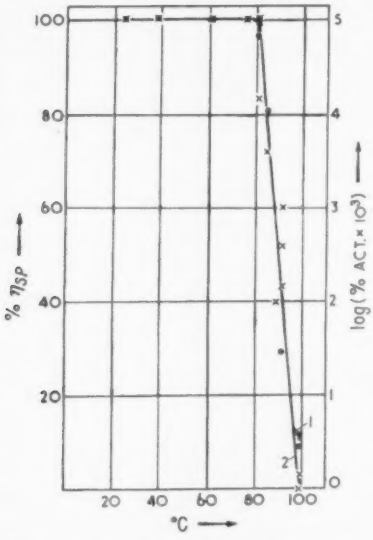


Fig. 1. Stability of the transforming principle preparation to heat. ●—viscosity (curve 1). ×—transforming activity (curve 2). The ordinates indicate the specific viscosities (all measured in standard buffer at 23°) as percentages of maximum viscosity, and the transforming activities as the logarithm of (percentage × 10^3) maximum activity; the abscissa indicates the temperature to which the sample was exposed for 1 hr prior to measuring viscosity and activity.

during the gradual removal of impurities (compare also[70]). HOTCHKISS[71] finds less than 0·02 per cent of protein in his purified transforming principle of pneumococcus. According to a recent estimate,[69] the amount of DNA sufficient to transform one cell of *H. influenzae* is of the order of 10^{-8} μg; an impurity of 0·01 per cent would correspond to about 6 molecules of a molecular weight of 10^5, or less than one of a molecular weight of 10^6. Thus one approaches the situation that the protein nature of the transforming principle can be excluded on purely analytical grounds.

Besides the above, recent studies offer some more suggestions (but not absolute proof) that the transforming principle is DNA. Crystalline pancreatic DNase in concentrations lower than 10^{-4} μg/cc produces a 10-fold decrease of activity within 20 minutes (Fig. 1[69]). Upon

heating, the temperature at which the transforming principle begins to lose its activity (81°, 1 hr) is the same as the temperature at which the

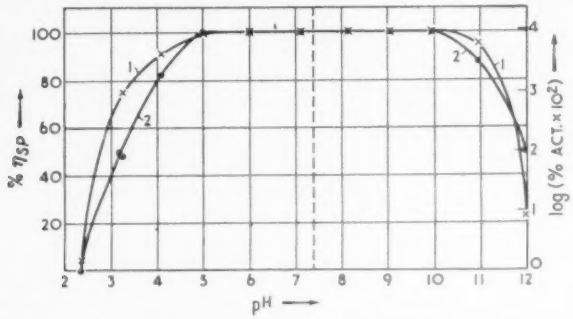


Fig. 2. Stability of the transforming principle preparation to *p*H changes. Curve 1 (×)—viscosity; curve 2 (●)—transforming activity. The ordinates indicate the specific viscosities as percentages of maximum viscosity, and the transforming activities as the logarithm of (percentage × 10^2) maximum activity; the abscissa indicates the *p*H to which the sample was exposed for 2 hr at 23°C.

viscosity of the bulk of the preparation begins to decrease (Fig. 2[69]). Most (known) proteins cannot withstand these heating conditions. The

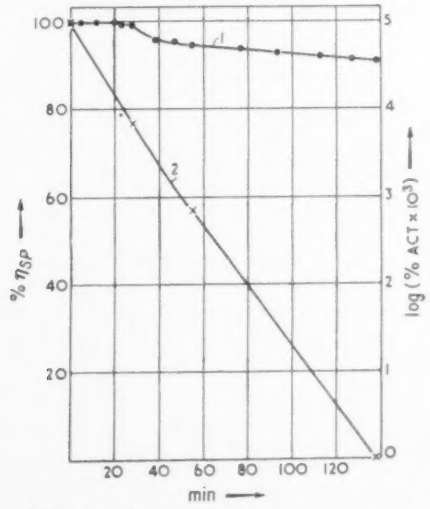


Fig. 3. Stability of the transforming principle preparation to crystalline pancreatic deoxyribonuclease. Curve 1 (●)—viscosities; curve 2 (×)—transforming activities. The ordinates as on Fig. 1; the abscissa indicates time of incubation with $8{\cdot}5 \times 10^{-5}$ μg/cc deoxyribonuclease at 30·1°.

*p*H values (on both acid and alkaline side) at which the activity begins to decrease are again the same as the *p*H values at which the viscosity

begins to decrease (Fig. 3[69]). Thus, the active molecule of the transforming principle seems to behave like the average molecule of DNA.

In summary, the extensive evidence favours the view that the transforming principle is DNA; no evidence to the contrary has ever been presented.

(*b*) *The chemical analysis*

As mentioned before, the chemical analysis of the purified transforming principles reveals only the presence of normal components of DNA: nitrogenous bases, deoxy sugar, and orthophosphate, all in the usual proportions.[7], [70], [68], [69] Of these, only the proportions of individual purines and pyrimidines may be of more than routine interest, in view of the now well recognized fact that these proportions (for the whole DNA) are characteristic for the species from which they were derived[72] (for a review, see[62]). Such analyses were thus far made only for the transforming principles of pneumococcus[70], [73] and of *Hemophilus influenzae*,[74], [75] as well as for the hypothetical transforming principle (DNA) of *Escherichia coli*.[76] These and other data are represented in Table 1. As can be seen from this Table, the molar ratios of individual

TABLE 1

Molar Ratios of Individual Purines and Pyrimidines in DNA of Man and of several Microorganisms

Species	$\frac{Adenine}{Guanine}$	$\frac{Thymine}{Cytosine}$
Man	1·6	1·7
Streptococcus pneumoniae .	1·45	1·75
Hemophilus influenzae .	1·75	1·54
Escherichia coli . .	1·04	0·95

purines and pyrimidines for the DNA of pneumococcus and of *H. influenzae*, although characteristics of their respective species, are not far from the values for the DNA of their host, man.[72] On the other hand, the values of DNA of *E. coli* are strikingly different; in fact, this is the only analyzed DNA in which the molar ratios of individual purines and pyrimidines are close to one. No satisfactory explanation of this peculiarity can be offered at present.

The principle that the composition of the whole DNA is characteristic of the species from which it was derived, does not necessarily imply that this composition is unchangeable: the alteration of DNA composition by change in environment has indeed been demonstrated.[77], [78], [79] When *E. coli* was grown in the presence of 5-bromouracil or 5-iodouracil, a considerable part of thymine in the DNA became replaced by these

VOL 6 195

unnatural thymine analogs. The change is reversible; in some strains, it also does not seem to affect the life of the cell.

(c) *The heterogeneity*

The data in Table 1 for each species refer to the whole DNA preparation from many, presumably identical cells. A chemist will, of course, be interested to know whether the preparation represents a single chemical species or a mixture of many. If one performs a transformation experiment using the principle from a donor carrying several transformable characters ("markers"), each of the resulting transformed cells carries, as a rule, only a single marker[32], [45], [39]; thus, the preparation behaves like a mixture of independent transforming particles. The doubly transformed cells occur as rarely as predictable from the probability of two independent particles hitting the same cell[45], [42] (an interesting exception will be discussed later on). On this basis one usually assumes that the DNA of each cell consists of many *different* molecules (different chemical species), each of which determines a different hereditary character.

The nature of the difference could be manifold. The molecules could differ in sequence of nucleotides or in steric pattern or in pattern of bonds, such as hydrogen bonds, or in some other unknown features of this kind; if these were the only differences, the analysis of the mixture could still represent the actual analysis of each molecule. If, on the other hand, the molecules differ in proportions of individual purines and pyrimidines then the analysis refers merely to the average composition. This difference in proportion could be due to an actual difference in pattern (which automatically means also some difference in sequence), or merely to the difference in length of the molecule: since the distribution of individual bases is dissymmetrical,[80] the difference in length may result in difference in proportion.

At present no method exists to test the possibility that the differences between various DNA molecules of the same species are merely differences in nucleotide sequence or bond pattern. On the other hand, the detection of the heterogeneity in length might be within the sensitivity of the present metnods. STERN and ATLAS[81] found that the DNA preparations from calf thymus on ultracentrifugation give polydispersity suggesting such an heterogeneity in length of the molecule. The solutions used for this study were rather dilute (0·04 per cent) as otherwise meaningless hyper-sharp boundaries are obtained. It must be stressed that such results are only as good as the preparation studied. Many steps in preparation, especially non-exclusion of the DNase action, may result in a depolymerized and therefore polydisperse artifact; indeed, the DNA prepared by an early HAMMARSTEN[82] method (DNase action not excluded) were polydisperse[81] even

at concentrations (0·9 per cent) at which carefully prepared DNA gives hypersharp boundaries.

That the DNA molecules of one species actually exhibit demonstrable differences in their proportions of individual purines and pyrimidines has been shown by the important discovery of DNA fractionation[83] (see also[84], [85], [86]). In these experiments it has been shown that the DNA preparations from calf thymus or from *E. coli* are non-homogeneous as far as their dissociation from basic proteins is concerned. This phenomenon has been used as a basis for separation of DNA into several fractions. This non-homogeneity would be explainable if the acidic characters of the molecules were different. Indeed, analysis of the fractions revealed that they differ in proportions of individual purines and pyrimidines. The compositions of the whole DNA preparations recorded in the literature are therefore not the compositions of individual molecules but of the averages of compositions.

It would be of obvious interest to study whether the transforming principle preparation carrying several "markers" can be separated into fractions, each containing one marker only. Attempts in this direction are under way in several laboratories, but to date no definite results have been published. One difficulty is that some fractionation procedures result in a considerable inactivation of the transforming principle so that the achieved "fractionation" may actually represent merely a "fractional inactivation."

(*d*) *The "native state" of DNA*

The study of the chemical nature and composition may eventually be undertaken even with degraded preparations of the transforming principle (provided that no fragments are lost). On the other hand, the physical study of the size and shape of the molecule demands that the subject of study be as "native" as possible.

At present, it is not known whether an even most carefully isolated DNA preparation can in all respects be identical with the DNA as it existed in the living cell[87]: the DNA is never found in the cell as such but merely as part of a nucleoprotein, which, in turn, may have bonds to still other substances. The term "native" when applied to the preparation of DNA *in vitro* is therefore rather misleading; what this usually means, is that such a preparation exhibits certain features which it must have had when it was in the living cell. The most obvious and the most important of these features is the proper biological activity. It is indeed a feature of which one can safely say that it could not have been created artificially by the process of isolation. Such an active DNA preparation may be called "*functionally intact*" and can serve as a yardstick: obviously, its properties are more constant and

are of more interest than the properties of the products of an undetermined degree of "denaturation"; in particular, the studies of the functionally intact DNA preparation may lead to correlation between the function and the structure.

As discussed before, the transforming activity is such a proper biological activity, which can be demonstrated in certain bacterial DNA preparations; these can therefore be called functionally intact. The studies of DNA preparations other than these are open to the criticism that the starting material was a product of degradation. While a completely satisfactory study of such DNA preparations cannot be offered at present, advantage can be taken of the fact that the DNA of different species exhibit similarities in their resistance to various agents[88]; thus, a DNA isolated under conditions which would not inactivate another DNA having transforming activity may lead to a "functionally intact" product. Another possibility is to add small amounts of active transforming principle during the isolation of any DNA.[88] This transforming principle is to serve as a marker of non-denaturation: the recovery of the activity in the final preparation of a DNA is an indication that the latter has not been mistreated.

VOL. 6 956

(*e*) *The isolation and purification of the transforming principles*

As discussed before, the methods for the isolation and purification of any transforming principle ought to be the same as for DNA, since the two appear thus far to be synonymous. However, the existing methods for isolation and purification of DNA were devised without the knowledge of the factors inactivating the transforming principle; some of the steps in these methods may indeed be harmless, but few of the published methods are acceptable as a whole.

AVERY, MACLEOD and MCCARTY[7] have made qualitative studies of some of the harmful factors; such studies have enabled them to devise a rational method of isolation and purification. They have realized that the action of DNase is by far more destructive than moderate heat (65°, 30 min.) and their method included heating the cells to destroy the enzyme before the cells are lysed. In a further study MCCARTY and AVERY[31] suggested an improved method in which the DNase action is inhibited by citrate during cell lysis; the citrate forms a non-ionized complex with Mg^{++} ion which is necessary for the action of *some* DNases. In both methods the cells were subsequently lysed by deoxycholate, the nucleic acid deproteinized by chloroform (SEVAG,[89]) the RNA removed by digestion with RNase, and the bacterial polysaccharide by digestion with specific enzyme[7], [31] (type III) or by fractional alcohol precipitation in presence of Ca^{++} ions (type II and VI[31], [90]). Most of the immunologically active substances of *H. influenzae* are not removable by the above methods. For this organism

another method has been devised[68] based on the observation that the DNA has higher electrophoretical mobility than the RNA or the immunologically active substances. In this method, as in the previous one, the cells are lysed with deoxycholate (in the presence of citrate, to inhibit DNase) and deproteinized; the DNA is subsequently separated in the ascending arm of a Tiselius electrophoretic cell. The method is more universal than the previous ones but less suitable for larger quantities of DNA. For this reason another method has been devised[69] in which the electrophoretic purification has been replaced by purification by fractional extraction. The method, applicable mainly to *H. influenzae*, is based on the observation that when the cells are lysed with deoxycholate and the lysate precipitated by alcohol, the impurities (RNA and immunologically active substances) are easily extractable from the precipitate whereas DNA only with difficulty. Accordingly, the precipitate is subjected up to 19 extractions, and the first extracts rejected. The last few extracts, upon additional deproteinization, consist almost entirely of DNA, with less than 0·4 per cent each of protein, RNA or immunologically active substances. The time-consuming deproteinization with chloroform was here replaced by a more efficient and simple deproteinization with sodium lauryl sulfate.[91], [92] The final product is stored at — 20° in solution in citrate buffer at *p*H 7; under these conditions the activity remains constant for at least a year.

The quantitative study of agents inactivating the transforming principle[69] has led to the realization of harmful effects of many procedures commonly used in the preparation of DNA. The most important harmful factors are listed below:

1. DNase of the cells from which DNA is being extracted. In some, but not all sources this enzyme can be inhibited prior to DNA extraction (by washing, heat, citrate, etc.); failure to do so inevitably results in an injured product.
2. Fe^{++} ion introduced by rusty tools and vessels.
3. Low ionic strengths (dialysis against water, dissolving in water).
4. Drying of final product, especially with alcohol and ether, and storage in the dried state.

All these factors will be discussed further on.

(*f*) *Physical properties of the transforming principle*

AVERY, MACLEOD and MCCARTY[7] have made some qualitative studies of the behaviour of the transforming principle on ultracentrifugation and on electrophoresis and concluded that in these respects the transforming principle resembles calf thymus DNA. The ultraviolet absorption spectrum was also similar to that of calf thymus DNA.

ZAMENHOF *et al.*[68] determined the electrophoretic mobility of the DNA of *H. influenzae* having transforming activity: the ascending mobility was -17×10^{-5} sq. cm/v/sec., in 0·15 M aqueous sodium chloride which was 0·02 M with respect to sodium phosphate (final *p*H 7·2). The entire transforming activity was associated with this DNA fraction. A quantitative study of the viscosity of a similar preparation has been also made;[69] the specific viscosity, when measured under similar conditions, was again similar to that of calf thymus DNA or human spleen DNA.[88]

An attempt to determine the molecular weight of pneumococcal transforming principle by electron and deuteron bombardment has been made by FLUKE, DREW and POLLARD:[93] the molecular weight so estimated was 6×10^6, which is similar to values obtained by the light scattering method for calf thymus DNA.[94], [95], [96]

An X-ray diffraction study of pneumococcal transforming principle has been made[97]; the crystalline structure was found to be the same as for calf thymus DNA.

In general, the physical properties of the transforming principles have not yet been extensively studied.

For studies of the physical properties of DNA having no biological activity, the reader is referred to previous reviews in volumes II and IV of this series.[63], [64]

(*g*) *The molecule of the transforming principle*

In the previous chapters each transforming principle was chiefly discussed as one substance; in the chapter on heterogeneity it was shown that the transforming principle preparation of a single species may be actually a mixture of substances. In this chapter the emphasis will be mainly on a single molecule of the transforming principle.

One of the first questions in such an approach is: how many molecules of transforming principle are necessary to transform one cell?

The lowest concentration of transforming principle which still accomplishes transformation in pneumococcus has been estimated by AVERY, MACLEOD and MCCARTY,[7] by MCCARTY and AVERY[31] and by HOTCHKISS.[70] The concentrations were of the order of 2×10^{-3} to 10^{-2} μg/ml; the weights referred either to the weights of a purified preparation, or the weights of DNA, which are equivalent for all practical purposes. Although seemingly low, these concentrations still represented 2×10^8 to 10^9 molecules of DNA (of molecular weight of the order of 5×10^6) per millilitre. In these studies the number of actually transformed cells was not counted and therefore the amount of DNA per one transformed cell could not be estimated. In general, for a study involving very small amounts of DNA, the pneumococcus is not best suited because of the DNase in the culture.

A quantitative study for the transforming principle of *H. influenzae* has been recently made by ZAMENHOF, ALEXANDER and LEIDY.[69] In one type of experiment a small total volume (0·02 ml) was used and the minimal total amount of DNA necessary to transform at least one cell was determined; the value so obtained was 3×10^{-7} μg. In another type of experiment the number of actually transformed cells was counted; the minimal amount of DNA per transformed cell was found to be 10^{-8} μg. Assuming the molecular weight of DNA to be of the order of 5×10^6, one can calculate that the number of molecules of transforming principle (computed as molecules of DNA) necessary to transform one cell in this system is of the order of 10^3.

It is of interest to compare this minimal amount of DNA with the total amount of DNA in one cell of *H. influenzae*. This latter amount has been found to be 2×10^{-9} μg[69]; thus the minimal amount of DNA found to transform one cell is only five times higher than the total amount of DNA in the cell. (This estimate is, of course, independent of the actual molecular weight of DNA.)

The chapter on heterogeneity of DNA contains a discussion of the premise that each DNA molecule of the cell is different because each determines a different hereditary character. If such an assumption is made here, then the number of molecules of DNA of *one kind* necessary for transformation would be of the order of five. If this number can further be reduced, support will be gained for the hypothesis that practically all molecules of one kind are active. If, in addition, every DNA molecule in the cell is assumed to be a potential transforming principle, then the physical and chemical behaviour of the bulk of the DNA preparation are representative of the physical and chemical behaviour of the active molecules. Obviously, more evidence is needed before such a view can be fully accepted, but even at the present status the "infectivity"[98] of DNA particles is comparable to that of bacterial viruses.

If one assumes the molecular weight of DNA to be of the order of 5×10^6, then the above figure of 2×10^{-9} μg DNA per cell corresponds to 250 molecules. Even if one assumes that all the molecules of DNA in the cell are functional, it still appears to be too few of them to serve as determinants of *all* hereditary characters of the cell. One is tempted to speculate that the DNA molecules do not determine *all* the characters and/or one molecule of DNA determines several characters. That the latter may be true is suggested by the discovery of the multiple transformations in *H. influenzae*[57] and subsequently in pneumococcus.[42] As mentioned before, in *H. influenzae* a new strain *ab* has been obtained by exposing cells *b* to the transforming principle from cells *a* (TP_a). The new type produces two capsules (*a* and *b*) at once and yields a new transforming principle TB_{ab} capable of producing cells *ab* from any

susceptible receptor cells. These results cannot be obtained by simply mixing TP_a and TP_b *in vitro*.

In pneumococcus, the exposure of sensitive cells to transforming principle (*MS*) from cells bearing two genetic markers, mannitol utilization (*M*) and streptomycin resistance (*S*), produces up to 15 times more cells bearing two markers (*MS*) than it would be expected from randomly distributed *independent* transformations. Again, this result cannot be obtained by simply mixing the transforming principles (*M*) and (*S*) *in vitro*.

Several explanations of these interesting phenomena appear possible at present. (1) Each of the two markers (*a* and *b* or *M* and *S*) is carried by a separate DNA molecule, but these molecules are linked by some such device as a protein bridge. The experimental evidence speaks against this explanation: the protein content of the preparation can be reduced to less than 0·02 per cent;[71] the usual deproteinization (involving protein denaturation[57], [42]) and autolytic proteolysis[42] did not abolish the "linkage." (2) The linkage between the two independent DNA molecules is caused by some direct bond which resists preparative methods; this explanation (purely speculative) would mean that the aggregate has double molecular weight, and that may result in some fractionation during the purification; no evidence of such a fractionation has been obtained. (3) If no evidence for explanation (1) or (2) will be produced, then the most likely explanation may be that, indeed, one molecule of DNA can determine more than one genetical marker. Just how one molecule of DNA can reproduce faultily in the presence of another one (in the same locus?) remains entirely in the domain of speculations. It is to be noted that the idea that the faulty reproduction of DNA (mutation?) is *always* an extremely rare phenomenon, may not be true: as mentioned before, the cells of *E. coli* grown in the presence of 5-bromouracil or 5-iodouracil, build these unnatural pyrimidines into the majority, if not all, molecules of DNA.[77], [78], [79]

(*h*) *The resistance to physical and chemical agents*

The literature on the effects of physical and chemical factors on DNA is rather voluminous; however, the starting material for these studies was in many cases prepared by the methods which are now known to give denatured and therefore less resistant DNA.

In this chapter, the discussion will be limited to studies in which the starting material had full transforming activity.

1. *Heat*

AVERY, MACLEOD and MCCARTY[7] reported that the crude (but not the purified) transforming principle of pneumococcus withstands heating

for 30 to 60 minutes, at 65°. ZAMENHOF, ALEXANDER and LEIDY[69] made a quantitative study of the resistance to heat of purified transforming principle of *H. influenzae* (Fig. 1, in section *a*, this chapter). From this figure it can be seen that both the viscosity and the activity of the transforming principle are indeed rather stable to heat: in citrate buffer the decrease (in both properties) occurs only after 1 hr of heating to temperatures higher than 81°. The change in concentration of DNA does not affect the stability. These stabilities are much greater than the above mentioned for the activity of the purified transforming principle of pneumococcus and for the viscosity of DNA of calf thymus.[90], [100] That this discrepancy is not due to the species' difference is shown by the fact that when the human DNA and calf thymus DNA were prepared and tested under conditions similar to those used for the transforming principle, they exhibited a similar stability of viscosity.[88] This high stability is undoubtedly partly due to the avoidance of unstabilizing steps (initial enzyme action, dialysis, drying). However, the possibility of a stabilizing effect of the citrate buffer cannot be excluded.[69]

As can further be seen from Fig. 1, the temperature coefficients of the (biological) inactivation rates between 85 and 95°C are very large, leading to large energies of activation. If the phenomenon for DNA is similar to thermal denaturation of protein, the above finding may mean that in order to produce comparable denaturation more bonds (H-bonds?) or stronger bonds must be broken in the case of DNA than in the case of most proteins.

It is interesting to note that those molecules of the transforming principle, which remained active after a sub-lethal treatment with heat, lose their stability to heating at 76° or even to storage in the cold.[101] This subject will be fully discussed in the last chapter.

For the hypotheses as to what may happen to the molecule of DNA on heating, the reader is referred to the literature.[99], [102], [103], [104]

2. *H^+ and OH^- ions*

AVERY, MACLEOD and MCCARTY[7] reported that the inactivation of pneumococcal transforming principle occurs rapidly at *p*H 5 and below. ZAMENHOF, ALEXANDER and LEIDY[69] made a quantitative study of the resistance to acid and alkali of purified transforming principle of *H. influenzae* (Fig. 2, in Section *a*, this chapter). From this figure it can be seen that both the viscosity and the activity are completely stable over a wide range symmetrical with respect to *p*H 7·4 (normal *p*H of the human blood). This stability may be species specific: it has been shown[87] that in yeast cells, whose cytoplasm has a much lower *p*H, the region of the stability of its DNA is extended towards a lower *p*H.

Beyond the stability range, a rapid drop of activity occurs accompanied by a loss of viscosity. At *p*H 5 (threshold of stability for DNA of *H. influenzae*) less than one primary phosphate group per 10^4 is undissociated; thus, one must conclude that in order to be active, the transforming principle must be in the form of a salt (deoxyribonucleate).

As in the case of heat, those molecules of the transforming principle which remained active after a sub-lethal treatment with acid lose their stability to heating at 76° (*p*H 7) or even to storage in the cold.[101]

The real mechanism of *irreversible* degradation outside the *p*H stability range is still an object of speculation. The reader is referred again to the literature dealing with the biophysical study on this matter for DNA in general.[105], [106], [95], [104] It is, however, to be added that for the low *p*H the inactivation may also be the result of removal of purines (and/or hypothetical hydrogen bonds to these purines). A study[69] of the amount of such a "depurination" of the transforming principle reveals that a 100-fold inactivation concurs with the removal of less than two purines per thousand; thus, practically every purine may be necessary for the activity.*

OL.
6
956

3. *Deoxyribonuclease*

AVERY, MACLEOD and MCCARTY[7] observed that crude DNases from dog intestinal mucosa, pneumococcus autolysates and normal sera inactivate the pneumococcal transforming principle. MCCARTY and AVERY[65] studied inactivation by purified pancreatic DNase, and HOTCHKISS[70] by streptococcal DNase and by crystalline pancreatic DNase; $2{\cdot}5 \times 10^{-4}$ μg of the latter per millilitre were sufficient to decrease the activity.

ZAMENHOF, ALEXANDER and LEIDY[69] made a quantitative study of the effects of crystalline pancreatic DNase on viscosity and on activity of the transforming principle of *H. influenzae* (Fig. 3). As can be seen from this figure, less than 10^{-4} μg of this enzyme per millilitre is sufficient to cause a ten-fold decrease in activity within 28 minutes, and a complete inactivation within 140 minutes (30°; 300 μg DNA/ml). On the other hand, the drop of viscosity at the beginning of inactivation is insignificant. A similar lag period in the depolymerizing action of the enzyme could be demonstrated when calf thymus DNA and human spleen DNA were used as substrates.[88] Despite this lag in depolymerization, the enzyme does exert some action in this period as evidenced by the marked decrease in activity[69] and by the loss of stability to heat[88], [101] of the DNA preparations exposed to such initial action of the enzyme. The nature of the initial changes is still not clear; they

* This may be true only on the assumption that the active molecule behaves like an average molecule of DNA.

may involve breaking of a few phosphate bonds insufficient to cause any decrease of size of the molecule still held together by the hydrogen bonds.

4. *Ionic strength*

AVERY, MACLEOD and MCCARTY[7] observed that the *purified* pneumococcal transforming principle rapidly loses activity if dissolved in distilled water. The quantitative study of the effects of exposure to various ionic strengths on activity and viscosity has been made for the transforming principle of *H. influenzae*.[69] Previous exposure to lower or higher ionic strengths did not affect the viscosities (as measured in a standard buffer). On the other hand, the activities were reduced by the exposure to lower ionic strength but not by higher ionic strength. The previous exposure to lower ionic strength was also found to decrease the stability to heat[88] (as measured in standard conditions). These results can be interpreted as follows:

It has long been known that the addition of salt to aqueous solutions of DNA decreases their viscosity.[82], [107], [108] It has been suggested[109], [110], [111] that the increase of viscosity in distilled water is due to actual increase in the length of the molecule, presumably by stretching (uncoiling) caused by repulsion of anions of the DNA molecule in the absence of salts. An inspection of the viscosity data[69] might suggest that this change is entirely reversible; however, the persistent loss of activity and stability to heat proves that this is not the case. The permanent damage may be due to breakage of a few vital bonds, such as hydrogen bonds, in a molecule rendered more vulnerable by distilled water;[100], [88], [104] the degree of thermal inactivation increases with temperature.[69]

5. *Dehydration*

AVERY, MACLEOD and MCCARTY[7] observed that drying of the purified pneumococcal transforming principle from the frozen state in the lyophile apparatus results in a loss of activity. Such a procedure usually involves dialysis prior to drying, and the reported inactivation may represent the combined effect of both these processes. The lyophilization has also been found to degrade calf thymus DNA.[88]

The quantitative study of the effect of various conditions of dehydration and storage on activity has been made for the transforming principle of *H. influenzae*.[69] None of the commonly used drying methods was harmless; the popular method of drying the DNA fibres with ethanol and ether resulted in an almost complete inactivation. Parallel studies on other DNA[88] revealed that the dried DNA loses its stability to heat (when tested in solution, in standard conditions); again, the most damaging method was drying with ethanol and ether.

The over-all effect of drying and storage in the dried state may actually be due to several causes: the instantaneous injury; slow injury by thermal oscillations in a molecule rendered more vulnerable; slow injury by progressive dehydration. The exact nature of the injuries is not known.

6. *Deamination*

The quantitative study of the effects of various purely chemical agents on activity was made only for the transforming principle of *H. influenzae*.[69]

Incubation of the transforming principle with 2 M or 1 M $NaNO_2$ at *p*H 5·3 resulted in a rapid inactivation of the transforming principle. However, the viscosity remained constant showing that the average DNA molecule was but slightly altered.

The deamination of adenine, guanine and cytosine in a high polymer is known to proceed slowly, even at a low *p*H.[112], [113] No studies of deamination at *p*H 5·3 have been reported, presumably because at this *p*H less than 1 per cent of free HNO_2 exists in an equilibrium with NO_2^-. An estimate was made of the extent of deamination corresponding to a 1000-fold decrease of activity: this extent was found to be of the order of 0·1 per cent. Thus it seems that practically all the primary amino groups must be intact for activity.*

7. *Mutagenic agents*

The agents in this very heterogeneous class will be grouped together here: although the mechanism of their action may be entirely different, they all seem to prime a phenomenon of great importance and of unknown nature, mutation.

It has recently been suggested[114] that the processes leading to mutation and to death are essentially the same, with the exception that the latter is accompanied by more extensive molecular changes. If the mutation indeed occurs through a direct action of the mutagenic agents on DNA, then the *inactivation* of transforming principle by these agents could be a demonstration of such a "too strong mutation."

It is to be emphasized that the transforming principle (DNA) *in vitro* can be harmlessly subjected to the action of a greater variety (and of more concentrated) mutagenic agents than the DNA in a living cell can; in the case of the transforming principle, the excess of such strong agents can be removed prior to the reintroduction of the transforming principle into the cell; in the case of the living cell, these strong agents may never reach the nuclear DNA, or they may cause death of the cell for reasons other than the reaction with DNA.

* This may be true only on the assumption that the active molecule behaves like an average molecule of DNA.

Of particular interest is the phenomenon of unstabilization of the *active* transforming principle as this may be the first step of the mutational process; this subject will be discussed in the last chapter.

Ultraviolet irradiation. The effect of U.V. on activity has been studied quantitatively for the transforming principle of *H. influenzae*.[115] A dose of 500 ergs/sq. mm was sufficient to cause 1000-fold inactivation. This dose is of the same order of magnitude as the one necessary to inactivate the bacterial viruses (for a review, see reference[116]). It is important to note that this dose is about 500 times lower than the lowest dose necessary to produce any demonstrable decrease of viscosity of DNA by U.V. irradiation.[117] For the extensive literature on the latter effect the reader is referred to the recent reviews[118], [64] in volumes III and IV of this series.

The U.V. irradiation also destroys the stability of the transforming principle to heat;[101] this subject will be discussed further on.

Ferrous ion and Hydrogen peroxide. The action of U.V. irradiation is now believed to be due to free radicals H· and OH· and peroxides formed during irradiation of organic molecules. Similar free radicals are believed to be produced in Fenton's reagent ($Fe^{++} + H_2O_2$) (for a review of these subjects see volumes III and IV of this series[118], [64]).

McCarty[119] observed reversible inactivation of pneumococcal transforming principle by ascorbic acid and several other self-oxidizing substances.

Ferrous ion self-oxidizes even in the absence of H_2O_2 and may produce free radicals. The quantitative study of the effects of Fe^{++} and H_2O_2, alone or together, on the activity has been made for the transforming principle of *H. influenzae*[69] (Table 2). H_2O_2 alone had a mild inactivating effect; on the other hand, Fe^{++} alone caused a 10-fold inactivation even in concentrations as low as 10^{-5} M* and complete inactivation in concentrations between $2{\cdot}8 \times 10^{-5}$ M and 2×10^{-4} M. The inactivation by Fe^{++} alone in concentrations up to 2×10^{-4} M is not accompanied by any change of viscosity of DNA; however, at concentrations ten times higher a rapid depolymerization occurs.[88]

The exact nature of the inactivating damage is still unknown; an oxidative deamination of a few nitrogeneous bases and/or breaking of a few vital labile bonds might be postulated. Extensive damage including depolymerization, deamination, dephosphorylation, splitting of bases and even breakage of sugar and of purine and pyrimidine rings can be produced by Fenton's reagents in concentrations much stronger than those used in the above mentioned study. The identification of the products of such strong reactions has recently been reported[120];

* The inactivation by minute traces of rust has been mentioned in the section on preparation of the transforming principle.

however, it is doubtful whether the same kind of products is involved in the mild reactions which cause inactivation of the transforming principle or which occur during sublethal mutagenic treatment by free radicals.

TABLE 2

Effects of Ferrous Ion and of Hydrogen Peroxide on Transforming Activity. Exposure for 3½ hr at 23°. DNA 314 μg/ml

Exp. No.	*Conc. of* Fe^{++} *M*	*Conc. of* H_2O_2 *M*	*Activity Units/μg DNA*
1	0	0	5000
2	0	$4{\cdot}4 \times 10^{-3}$	500
3	0	3×10^{-4}	5000
4	2×10^{-7}	3×10^{-4}	5000
5	2×10^{-6}	3×10^{-4}	5000
6	3×10^{-6}	0	5000
7	$1{\cdot}25 \times 10^{-5}$	0	500
8	2×10^{-5}	3×10^{-4}	100
9	2×10^{-5}	0	100
10	$2{\cdot}8 \times 10^{-5}$	0	50
11	2×10^{-4}	3×10^{-4}	0
12	2×10^{-4}	0	0
13	2×10^{-4}	$4{\cdot}4 \times 10^{-3}$	0
14	$1{\cdot}3 \times 10^{-4}$	3×10^{-4}	0

Formaldehyde, reported to be a mutagenic agent,[121], [122] in strong concentrations (4 M) causes both inactivation and decrease of viscosity of the transforming principle of *H. influenzae*.[69] A probable interpretation of these findings is that formaldehyde reacts slowly with the primary amino groups of adenine, guanine and/or cytosine, and that in this reaction enough labile bonds (hydrogen bonds?) are broken to cause a decrease of asymmetry of the molecule as well as inactivation; the latter may again be indicative of the importance of free amino groups for the activity of the transforming principle.

Mustards. HERRIOTT[123] studied inactivation of pneumococcal transforming principle by di(2-chloroethyl) sulfide (mustard gas). This mustard in concentration as low as 6×10^{-5} M was able to cause

complete inactivation in 2 hr. The inactivation of the transforming principle of *H. influenzae* by various nitrogen mustards was studied by ZAMENHOF *et al.*[115] Di(2-chloroethyl)-methylamine (HN_2) causes a complete inactivation at a mustard concentration of 10^{-4} M. 10-fold decrease of activity is observed at a mustard concentration of 10^{-5} M (6 hr., 23°) which is 300 times lower than the lowest concentration producing any detectable decrease of viscosity of DNA.[117] For this and other nitrogen mustards the order of inactivation power for the transforming principle seems to be the same as the order of carcinostatic power of these compounds, thus suggesting a correlation between these phenomena.[115] Such results cannot be predicted from the chemical reactivities of the compounds. The transforming principle which survives the mustard treatment also undergoes a change; it becomes unstable to heat.[101]

The conceivable reactions of mustards and DNA are: (1) alkylation of amino groups (especially primary amino groups) of nitrogenous bases; (2) ether formation on enolic hydroxyls of nitrogenous bases; and (3) esterification of the phosphate groups. At present it is not known which of these reactions predominates, or is responsible for, inactivation or unstabilization at the low concentrations of mustards sufficient for these purposes.

Non-inactivating agents. The study of agents which even in high concentrations do not inactivate the transforming principles may be of interest both from the theoretical and the practical (preparative) point of view. Such a quantitative study has been made for the transforming principle of *H. influenzae*.[69], [115] For convenience only, these agents have been divided into groups:

(1) *Protein denaturing and sterilizing agents* in many cases do not cause any inactivation of the transforming principle. This finding may serve as a further indication that the transforming principle is not a protein or a virus; and that to produce denaturation more bonds (H-bonds?) or stronger bonds must be broken in the case of DNA than in the case of most protein (see also section "Heat"). From the practical point of view, such a study helps to devise deproteinization methods harmless for DNA.

(2) *Agents reported to have mutagenic action*, in general, strongly inactivate the transforming principle; however, a few were found (manganous ions, urethane, acriflavine, etc.) to have no inactivating effect at all. The nature of this discrepancy is not clear; one of many possible explanations is that such agents need the presence of the cell to show any action, either mutagenic or lethal.

(3) *Carcinogenic and carcinostatic agents* can be divided into two classes:[115] those (as discussed before) which even in minute doses completely inactivate the transforming principle (U.V., mustards);

and those which have no inactivating effect at all (methylcholanthrene, urethane, carcinogenic azo dyes). The mechanism of action of some of these agents (such as azo dyes) may never involve DNA at all; others may again need the presence of the cell to show any action.

8. *The unstable transforming principle*

As mentioned above, the sublethal treatment of the transforming principle of *H. influenzae* by heat, acids, mustards, deoxyribonuclease and U.V. irradiation* results in a product which is unstable to heat or even to storage in the cold.[101] The study of the kinetics of inactivation by heat reveals that more than one reaction is involved (such as unstabilization followed by inactivation). When the transforming principle, rendered very unstable by a sublethal heat or mustard treatment, was used for transformation experiments, the transformed cells yielded transforming principle which was completely stable; thus, the change (unstabilization) *in vitro* was not retained on reproduction and therefore was not a "mutation *in vitro*."

The chemical nature of a true mutation is, of course, still unknown; if it concerns DNA at all, it may involve any of the changes mentioned in the discussion of differences between individual DNA molecules (sequence of nucleotides, proportions, etc.) (see chapter "Heterogeneity"). At present, any such change can easiest be visualized as a (rare) fault occurring during DNA reproduction. However, the probability and the location of the fault along the DNA molecule may be determined by a preceding process, the gene unstabilization.[126]

It has repeatedly been suggested[127], [128], [129], [130] that the first action of the mutagenic agents is to bring the gene into an unstable state from which it can either return to the previous stable state or change into a stable mutant gene; this secondary change may not require the presence of the mutagenic agent any more. The phenomenon of delayed mutations may point in this direction. It is conceivable that such unstabilization is also possible *in vitro* and that the unstabilization of the transforming principle is indeed a demonstration of such a change.

ACKNOWLEDGMENTS

This review was undertaken as part of the research projects supported by grants from the National Institute of Health, Public Health Service, by a grant-in-aid from the American Cancer Society upon recommendation of the Committee on Growth of the National Research Council, and by an institutional grant from the American Cancer Society to Columbia University.

* The U.V.-induced instability to heat is not limited to DNA: it has been also reported for viruses and enzymes.[124], [125]

References

(1) McCarty, M.; *Bact. Reviews* 10 (1946) 63
(2) Austrian, R.; *Bact. Reviews* 16 (1952) 31
(3) Hotchkiss, R. D.; *Exp. Cell. Res.*, suppl. 2 (1952) 383
(4) Griffith, F.; *J. Hyg.* 27 (1928) 113
(5) Dawson, M. H. and Sia, R. H. P.; *J. Exp. Med.* 54 (1931) 681
(6) Alloway, J. L.; *J. Exp. Med.* 55 (1932) 91; 57 (1933) 265
(7) Avery, O. T., MacLeod, C. M. and McCarty, M.; *J. Exp. Med.* 79 (1944) 137
(8) Berry, G. P. and Dedrick, H. M.; *J. Bact.* 31 (1936) 50
(9) Berry, G. P.; *Arch. Path.* 24 (1937) 533
(10) Hurst, E. W.; *Brit. J. Exp. Path.* 18 (1937) 23
(11) Hoffstadt, R. E. and Pilcher, K. S.; *J. Infect. Dis.* 68 (1941) 67
(12) Gardner, R. E. and Hyde, R. R.; *J. Infect. Dis.* 71 (1942) 47
(13) Houlihan, R. B.; *Proc. Soc. Exp. Biol. Med.* 51 (1942) 259
(14) Boivin, A., Vendrely, R. and Lehoult, Y.; *C. R. Acad. Sci. Paris* 221 (1945) 646
(15) Boivin, A.; *Cold Spring Harbor Symp. Quant. Biol.* 12 (1947) 7
(16) Boivin, A., Tulasne, R. and Vendrely, R.; *C. R. Acad. Sci. Paris* 225 (1947) 703
(17) Atwood, K. C., Schneider, L. K. and Ryan, F. J.; *Cold Spring Harbor Symp. Quant. Biol.* 16 (1951) 345
(18) Zamenhof, S.; *J. Genetics* 47 (1945) 69
(19) Weil, A. J. and Binder, M.; *Proc. Soc. Exp. Biol. Med.* 66 (1947) 349
(20) Dianzani, M. U.; *Boll. ist. Sieroterap. Milan.* 29 (1950) 161; *Experientia* 6 (1950) 332; *Exp. Cell Res.* 5 (1953) 311
(21) Roland, F. and Stuart, C. A.; *Antibiotics and Chemotherapy* 1 (1951) 523
(22) Katsunuma, N. and Nakasato, H.; *Kekkaku* 29 (1954) 19; (*Chem. Abstr.* 48 (1954) 5281)
(23) Coleman, M. F. and Reid, J. J.; *Phytopathol.* 39 (1949) 182
(24) Klein, D. T. and Klein, R. M.; *J. Bact.* 66 (1953) 220
(25) Braun, W., Whallon, J. and Mauzy, W. L.; *Records Gen. Soc. Amer.* 23 (1954) 35
(26) Alexander, H. E. and Leidy, G.; *Proc. Soc. Exp. Biol. Med.* 73 (1950) 485; *J. Exp. Med.* 93 (1951) 345
(27) Alexander, H. E. and Redman, W.; *J. Exp. Med.* 97 (1953) 797
(28) Tunevall, G.; *Acta Pathol. Microbiol. Scand.* 31 (1952) 233
(29) Leidy, G.; Unpublished experiments
(30) Chargaff, E., Zamenhof, S., Brawerman, G. and Kerin, L.; *J. Am. Chem. Soc.* 72 (1950) 3825
(31) McCarty, M. and Avery, O. T.; *J. Exp. Med.* 83 (1946) 97
(32) Austrian, R. and MacLeod, C. M.; *J. Exp. Med.* 89 (1949) 439, 451
(33) Zamenhof, S., Leidy, G., Fitzgerald, P. L., Alexander, H. E. and Chargaff, E.; *Fed. Proc.* 11 (1952) 315; *J. Biol. Chem.* 203 (1953) 695
(34) Zamenhof, S. and Leidy, G.; *Fed. Proc.* 13 (1954) 327
(35) McPherson, C. F. C., Heidelberger, M., Alexander, H. E. and Leidy, G.; *J. Immunol.* 52 (1946) 207
(36) Zamenhof, S. and Leidy, G.; Unpublished results
(37) MacLeod, C. M. and Krauss, M. R.; *J. Exp. Med.* 86 (1947) 439
(38) Taylor, H. E.; *C. R.* 228 (1949) 1258; *J. Exp. Med.* 89 (1949) 399
(39) Austrian, R.; *J. Exp. Med.* 98 (1953) 21; 98 (1953) 35
(40) Ephrussi-Taylor, H.; *Exp. Cell Res.* 6 (1954) 94
(41) Austrian, R. and Colowick, M.; *Bull. Johns Hopkins Hosp.* 92 (1953) 375

VOL
6
195

REFERENCES

(42) HOTCHKISS, R. D. and MARMUR, J.; *Proc. Natl. Acad. Sci.* 40 (1954) 55
(43) MARMUR, J. and HOTCHKISS, R. D.; in preparation
(44) HOTCHKISS, R. D.; *Harvey Lectures* 49 (1953–54) 124
(45) HOTCHKISS, R. D.; *Cold Spring Harbor Symp. Quant. Biol.* 16 (1951) 457; *Symposium sur le mode d'action des antibiotiques, IIe Congrès Int. de Biochimie* (1952) 40, Paris, Société d'Edition d'Enseignement Supérieur
(46) DEMEREC, M.; *Proc. Natl. Acad. Sci.* 31 (1945) 16
(47) ALEXANDER, H. E. and LEIDY, G.; *J. Exp. Med.* 97 (1953) 17
(48) ALEXANDER, H. E. and LEIDY, G.; *Proc. Soc. Exp. Biol. Med.* 78 (1951) 625
(49) MCCARTY, M., TAYLOR, H. E. and AVERY, O. T.; *Cold Spring Harbor Symp. Quant. Biol.* 11 (1946) 177
(50) HOTCHKISS, R. D. and EPHRUSSI-TAYLOR, H.; *Fed. Proc.* 10 (1951) 200
(51) RAVIN, A. W.; *Exp. Cell Res.* 7 (1954) 58
(52) ALEXANDER, H. E., LEIDY, G. and HAHN, E.; *J. Exp. Med.* 99 (1954) 505
(53) HOTCHKISS, R. D.; *Proc. Natl. Acad. Sci.* 40 (1954) 49
(54) DE LAMATER, E. D.; *Cold Spring Harbor Symp. Quant. Biol.* 16 (1951) 381
(55) ROWEN, J. W. and NORMAN, A.; *Arch. Biochem. Biophys.* 51 (1954) 524
(56) EPHRUSSI-TAYLOR, H.; *Exp. Cell Res.* 2 (1951) 589
(57) LEIDY, G., HAHN, E. and ALEXANDER, H. E.; *J. Exp. Med.* 97 (1953) 467
(58) LEDERBERG, J., LEDERBERG, E., ZINDER, N. D. and LIVELY, E. R.; *Cold Spring Harbor Symp. Quant. Biol.* 16 (1951) 413
(59) ZINDER, N. D. and LEDERBERG, J.; *J. Bact.* 64 952) 679
(60) ZINDER, N. D.; *Cold Spring Harbor Symp. Quant. Biol.* 18 (1953) 261
(61) DEMEREC, M.; personal communication
(62) ZAMENHOF, S. in *Phosphorus Metabolism* 2 (1952) 301, Baltimore, The Johns Hopkins Press
(63) JORDAN, D. O.; *Progress in Biophysics* 2 (1951) 51
(64) DAVIDSON, P. F., CONWAY, B. E. and BUTLER, J. A. V.; *Progress in Biophysics* 4 (1954) 148
(65) MCCARTY, M. and AVERY, O. T.; *J. Exp. Med.* 83 (1946) 89
(66) LEVENE, P. A. and BASS, L. W.; *Nucleic acids* 1931, New York, The Chemical Catalog Co.
(67) MIRSKY, A. E.; *Cold Spring Harbor Symp. Quant. Biol.* 12 (1947) 15
(68) ZAMENHOF, S., LEIDY, G., ALEXANDER, H. E., FITZGERALD, P. L. and CHARGAFF, E.; *Arch. Biochem. Biophys.* 40 (1952) 50
(69) ZAMENHOF, S., ALEXANDER, H. E. and LEIDY, G.; *J. Exp. Med.* 98 (1953) 373
(70) HOTCHKISS, R. D.; *Coll. int. Cent. natl. recherche Sci., Unités biol. douées contin. génét.* 8 (1949) 57
(71) HOTCHKISS, R. D. in *Phosphorus Metabolism* 2 (1952) 426, Baltimore, The Johns Hopkins Press
(72) CHARGAFF, E., ZAMENHOF, S. and GREEN, C.; *Nature*, Lond. 165 (1950) 756
(73) DALY, M. M., ALLFREY, V. G. and MIRSKY, A. E.; *J. Gen. Physiol.* 33 (1949–50) 497
(74) ZAMENHOF, S.; *Bull. N.Y. Acad. Med.* 28 (1952) 349
(75) ZAMENHOF, S., BRAWERMAN, G. and CHARGAFF, E.; *Biochem. Biophys. Acta* 9 (1952) 402
(76) GANDELMAN, B., ZAMENHOF, S. and CHARGAFF, E.; *Biochim. Biophys. Acta* 9 (1952) 399
(77) DUNN, D. B. and SMITH, J. D.; *Nature*, Lond. 174 (1954) 305
(78) ZAMENHOF, S. and GRIBOFF, G.; *Nature*, Lond. 174 (1954) 306; 307
(79) ZAMENHOF, S.; *Abstracts. Am. Chem. Soc.* 126th meeting, New York (1954) 42C

(80) ZAMENHOF, S. and CHARGAFF, E.; *J. Biol. Chem.* 178 (1949) 531; 187 (1950) 1
(81) STERN, K. G. and ATLAS, S. M.; *J. Biol. Chem.* 203 (1953) 795
(82) HAMMARSTEN, E.; *Biochem. Z.* 144 (1924) 383
(83) CHARGAFF, E., CRAMPTON, C. F. and LIPSHITZ, R.; *Nature*, Lond. 172 (1953) 289
(84) BROWN, G. L. and WATSON, M.; *Nature*, Lond. 172 (1953) 339
(85) LUCY, J. A. and BUTLER, J. A. V.; *Nature*, Lond. 174 (1954) 32
(86) BENDICH, A.; *Abstracts, Am. Chem. Soc.* 126th meeting, New York (1954) 52C
(87) ZAMENHOF, S. and CHARGAFF, E.; *J. Biol. Chem.* 186 (1950) 207
(88) ZAMENHOF, S., GRIBOFF, G. and MARULLO, S.; *Biochim. Biophys. Acta* 13 (1954) 459
(89) SEVAG, M. G.; *Biochem. Z.* 273 (1934) 419
(90) KOENIG, V. L., LARKINS, L. and PERRINGS, J. D.; *Arch. Biochem. Biophys.* 39 (1952) 355
(91) MARKO, A. M. and BUTLER, G. C.; *J. Biol. Chem.* 190 (1951) 165
(92) DOUNCE, A. L., SIMMONS, N. S. and KAY, E. R. M.; *Fed. Proc.* 10 (1951) 177
(93) FLUKE, D., DREW, R. and POLLARD, E.; *Proc. Natl. Acad. Sci.* 38 (1952) 180
(94) DOTY, P. and BUNCE, B. H.; *J. Am. Chem. Soc.* 74 (1952) 5029
(95) ROWEN, J. W.; *Biochim. Biophys. Acta* 10 (1953) 391
(96) REICHMANN, M. E., RICE, S. A., THOMAS, C. A. and DOTY, P.; *J. Am. Chem. Soc.* 76 (1954) 3047
(97) WILKINS, M. H. F., STOKES, A. R. and WILSON, H. R.; *Nature*, Lond. 171 (1953) 738
(98) HOTCHKISS, R.; in *Dynamics of virus and rickettsial infections*, (1954) 405, New York, Blakiston
(99) GOLDSTEIN, G. and STERN, K. G.; *J. Polymer. Sc.* 5 (1950) 687
(100) MIYAJI, T. and PRICE, V. E.; *Proc. Soc. Exp. Biol. Med.* 75 (1950) 311
(101) ZAMENHOF, S., LEIDY, G. and HAHN, E.; *Records Genetics Soc. Amer.* 23 (1954) 75
(102) COSGROVE, D. J. and JORDAN, D. O.; *Tr. Faraday Soc.* 46 (1950) 793
(103) SHACK, J. and THOMPSETT, J. M.; *J. Biol. Chem.* 197 (1952) 17
(104) THOMAS, R.; *Biochim. Biophys. Acta* 14 (1954) 231
(105) SCHWANDER, H.; *Helvet. Chim. Acta* 32 (1949) 2510
(106) REICHMANN, M. E., BUNCE, B. H. and DOTY, P.; *J. Polymer Sci.* 10 (1953) 109
(107) GREENSTEIN, J. P. and JENRETTE, W. V.; *Cold Spring Harbor Symp. Quant. Biol.* 9 (1941) 236
(108) CREETH, J. M., GULLAND, J. M. and JORDAN, D. O.; *J. Chem. Soc.* (1947) 1141
(109) SMITH, D. B. and SHEFFER, H.; *Canad. J. Research*, Section B, 28 (1950) 96
(110) JORDAN, D. O.; *Tr. Faraday Soc.* 46 (1950) 792
(111) BASU, S.; *Nature*, Lond. 168 (1951) 341; *Science* 115 (1952) 465
(112) BREDERECK, H., KÖTHNIG, M. and LEHMANN, G.; *Ber. chem. Ges.* 71 (1938) 2613
(113) FLETCHER, W. E., GULLAND, J. M., JORDAN, D. O. and DIBBEN, H. E.; *J. Chem. Soc.* (1944) 30
(114) MCELROY, W. D.; *Science* 115 (1952) 623
(115) ZAMENHOF, S., LEIDY, G. and REINER, B.; *Proc. Amer. Assn. Cancer Res.* 1 (1954) 53
(116) LURIA, S. E.; *General Virology*, 1953, New York, J. Wiley & Sons
(117) ZAMENHOF, S.; in preparation

REFERENCES

(118) ERRERA, M.; *Progress in Biophysics* 3 (1953) 88
(119) MCCARTY, M.; *J. Exp. Med.* 81 (1945) 501
(120) SCHOLES, G. and WEISS, J.; *Biochem. J.* 53 (1953) 567
(121) RAPOPORT, I. A.; *C. R. Acad. Sc.*, U.R.S.S. 54 (1946) 65
(122) ENGLESBERG, E.; *J. Bact.* 63 (1952) 1
(123) HERRIOTT, R. M.; *J. Gen. Physiol.* 32 (1948) 221
(124) KLECZKOWSKI, J. and KLECZKOWSKI, A.; *J. Gen. Microbiol.* 8 (1953) 135
(125) KLECZKOWSKI, A.; *Biochem. J.* 56 (1954) 345
(126) ZAMENHOF, S.; *Science* 120 (1954) 791
(127) AUERBACH, C., ROBSON, J. M. and CARR, J. G.; *Science* 105 (1947) 243
(128) WITKIN, E. M.; *Cold Spring Harbor Symp. Quant. Biol.* 16 (1951) 357
(129) MCELROY, W. D. and SWANSON, C. P.; *Quarterly Review Biol.* 26 (1951) 348
(130) KAPLAN, R. W.; *Z. Naturforsch.* 7b (1952) 291

VOL.
6
1956

VOL
6
195

5

BIOPHYSICAL ASPECTS OF NEURO-MUSCULAR TRANSMISSION

J. del Castillo and B. Katz

CONTENTS

		Page
I.	INTRODUCTION	122
II.	*A.* ELECTRICAL EXCITATION AND PROPAGATION OF IMPULSES IN NERVE AND MUSCLE FIBRES	123
	B. CHEMICAL TRANSMISSION FROM MOTOR NERVE ENDINGS TO MUSCLE END-PLATE	127
III.	CRITICISMS OF THE CHEMICAL TRANSMISSION THEORY	128
	A. Does Electrotonic Transmission take place across the Neuro-muscular Junction?	129
	B. Are the Quantities of Acetylcholine released from Nerve Endings inadequate to stimulate Motor End-plates?	131
	C. On the Effect of Intra-cellular Application of Acetylcholine and Related Substances	133
IV.	THE DIFFERENT STAGES OF NEURO-MUSCULAR TRANSMISSION	137
	A. The Release of Acetylcholine from Motor Nerve-endings	137
	1. The Role of Inorganic Ions in the Release of Acetylcholine	139
	(*a*) Potassium	139
	(*b*) Sodium	139
	(*c*) Calcium and Magnesium	140
	2. The Spontaneous Quantal Release of Acetylcholine	141
	3. The Relation between "Responsive" and "Spontaneous" Release of ACh	144
	4. Other Factors Controlling the Release of ACh	146
	B. The Mechanism of the ACh/Receptor Action	150
	1. Transfer of ACh from Pre- to Post-synaptic Membrane	150
	2. The Electromotive Effect of Acetylcholine	151
	Does ACh act by inducing a Permeability Change or by specific Penetration of the Membrane?	152
	3. The Nature of the End-plate Potential	154
	C. The Removal of the Transmitter from its Site of Action	164
V.	CONCLUSIONS	166
	REFERENCES	167

5

BIOPHYSICAL ASPECTS OF NEURO-MUSCULAR TRANSMISSION

J. del Castillo and B. Katz

I. Introduction

The study of physical events which occur during excitation and the transmission of impulses in nerve and muscle cells has been greatly promoted in recent years by the development of more direct electric recording techniques. It is now possible with the help of an internal electrode to measure the potential difference across the surface membrane of a single cell and to obtain faithful records of its changes during the passage of an excitation wave.[1–6] The introduction of this method was followed by a rapid advance in our knowledge of the mechanism of the nerve impulse. The same method has been applied to problems of inter-cellular, i.e. "synaptic" transmission[7–10], but the situation at a synapse is technically not as favourable as in a large single cell. In general, the pre-synaptic nerve endings are unmanageably small, and practically all the electrical evidence which relates to events at cell contacts has been obtained from the large post-synaptic structures, e.g. the end-plate region of the muscle fibre, or the motoneurone in the spinal cord.

The purpose of this article is to review recent experimental studies of nerve-muscle transmission and to discuss new concepts which have arisen as a result of this work during the last five years. The review is confined mainly to the biophysical problems inherent in the transmission of a single impulse from a motor nerve axon to a skeletal muscle fibre. It shares some ground with a survey recently made by Fatt[11] in which problems of synaptic excitation and inhibition are dealt with more generally.* An important adjoining field, viz. the relation between chemical structure and pharmacological action of neuro-muscular stimulants and blocking agents has been well covered by Riker[12] (see also [13–15]), and the previous position, up to 1950, has been reviewed by Hunt and Kuffler[16] and by Rosenblueth.[17]

Until quite recently, there has been recurrent controversy as to whether a fundamental distinction really exists between the transmission of excitatory events in single cells, e.g. impulses in nerve or muscle fibres, and the transmission across cell contacts, e.g. at the

* There is a further restriction in that this article is concerned only with the classical "tritch-system." The special problems of the "slow-fibre system" of the frog[10a] have not been considered here.

nerve-muscle junction.[18], [19] It will be of interest, therefore, to compare the two processes and to view recently acquired knowledge on the myoneural synapse against the background of existing information on the axon membrane.

II

A. Electrical excitation and propagation of impulses in nerve and muscle fibres

Excitation in nerve or muscle is due to a sudden change in the fibre membrane by which electrical energy is released in an "explosive" manner. The transition from the unexcited (or "resting") to the excited (or "active") state occurs when the potential difference which normally exists across the fibre membrane is rapidly reduced to a critical level (the "threshold"). At this point, the potential change becomes self-reinforcing and undergoes a rapid automatic cycle, known as the "spike" or "action potential." Some quantities are shown in Fig. 1 for the case of a large non-medullated axon (see HODGKIN[20]): the "resting potential" (i.e. p.d. across the unstimulated fibre membrane) is about 60 mV, inside negative, the "threshold potential" is about 45 mV (hence the "threshold depolarization," that is the reduction of the resting potential needed to activate the fibre, is 15 mV). The amplitude of the action potential is about 100 mV; the p.d. across the membrane at the peak of activity (the so-called "active-membrane potential") is 40 mV, inside positive. Thus, during activity, the membrane potential reverses sign, and its total swing (*a–c*) is several times greater than the threshold change (*a–b*) initially required to excite.

The electrical changes are brought about by movement of ions across the axon surface, and the key to an understanding of the action potential was provided by the discovery that the conductances (or "permeabilities") of the fibre membrane to different cations, particularly to sodium and potassium, are highly specific and depend critically on the level of the membrane potential.[21] The immediate effect of a depolarization, for instance, is a specific increase in sodium permeability.

The "resting" fibre is in a steady state of metabolic activity whose free energy is, in some as yet unknown way, utilized to preserve large concentration gradients of Na and K across the fibre surface (see HODGKIN and KEYNES[22]). The cell contains 20–50 times more potassium and 3–15 times less sodium than the external fluid.[20] These concentration differences are maintained *in vivo*, in spite of a steady slow turnover of both kinds of ions between inside and outside. Studies with radio-isotopes have shown that the rate of penetration of ions through the resting fibre surface is very slow, and that of sodium much slower than of potassium.[23], [24] On the other hand, once the labelled

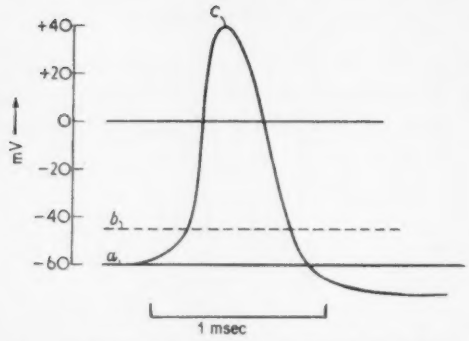


Fig. 1*a*. This diagram shows a membrane action potential from a giant nerve fibre of the squid, at about 20°C. The potential difference is recorded between the interior of the cell and the surrounding fluid. (*a*) resting potential, (*b*) threshold level (see Fig. 1*b*), (*c*) peak potential during activity.

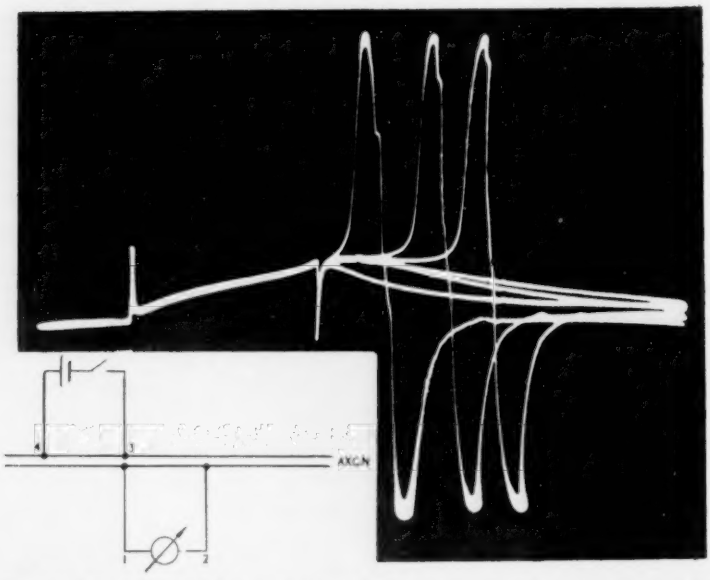


Fig. 1*b*. The "threshold of excitation." These records illustrate the unstable state which occurs when the membrane potential is displaced from the resting to the threshold level. Six superimposed records obtained from a crab nerve fibre with a pair of external electrodes, 1 and 2 (KATZ, unpublished records, reproduced by permission of I.C.I. Film Unit; cf. 119, 120). The axon was stimulated through a separate pair of surface electrodes, 3 and 4 with cathode at 3, by a current pulse (duration about 2 msec., beginning and end marked by vertical artifacts). The recording electrode 1 was placed near the cathode. When the stimulus exceeded threshold, a propagating (hence diphasic) action potential arose. The current was adjusted to, and kept constant at, the critical intensity at which propagated spikes were elicited at random in about half the trials. The effect of the current pulse was to change the membrane potential at the cathode to a level at which it was unstable. At this "threshold potential," the electrical events are left in a state of temporary balance, for a variable length of time, between regenerative and restorative changes (now identified respectively, with inward Na and outward K currents[20].)

potassium ions had entered the cytoplasm they appeared to move freely within the cell boundaries, indicating that apart from its surface layer, the protoplasmic contents of the cell offer no special resistance to the diffusion of these ions.[25], [26] Electrical measurements[27–29] have confirmed that the fibre membrane has an extremely high specific resistance, while the conductivity of the axoplasm is of the same order of magnitude, though not quite as high, as that of the outside fluid.

The resting membrane is relatively much more permeable to K than to Na, and this would explain the direction and approximate magnitude of the resting potential. When an electric current is passed *outward* through the fibre membrane, it reduces the resting potential, and as a result the specific sodium conductance increases. The basis of this effect is unknown, but its consequences are extremely important. Na ions will now flow at an increased rate, from the higher outside level of concentration and electric potential into the fibre and by transferring positive charge into the interior further reduce the resting potential. This leads to a reinforcement of the initial depolarization and so to an additional increase of Na-permeability. At a certain level, the process becomes "explosive" and an action potential arises. It will be noted that the entry of Na ions takes place in apparent opposition to Ohm's law, and the properties of the axon membrane in respect to sodium show the characteristics of a negative resistance, allowing more current to flow as the p.d. is reduced. This regenerative interaction between Na-influx and Na-permeability is the basis of electric excitation. Entry of sodium into the fibre would continue until an equilibrium is reached, at which the inward concentration gradient of Na ions is balanced and their further net entry prevented by an outward directed potential gradient, the reverse of the resting potential. This explains why the membrane potential is reversed during the ascent of the spike. The full attainment, however, of a Na-equilibrium potential is prevented by further events: the work of HODGKIN and HUXLEY[21] has shown that the specific increase of Na-permeability is of a transient nature, even when the membrane potential is artificially held at a depolarized level. After a delay of one or a few milliseconds, the condition of specifically high Na-conductance is terminated and converted into a state of specifically high K-conductance.[30]

To sum up, both Na- and K-permeabilities increase when the membrane is depolarized, but the Na change precedes in time and is transient, while the K change is delayed and maintained. The important point is that the two specific permeability changes have different time lags and are out of phase. In the normal course of events, the rise of the action potential is brought about by rapid entry of Na into the fibre, while accelerated outflow of K serves to re-establish the resting condition and to ensure a quick decline of the action potential, followed

by a slower recovery of the normal membrane conductances. It will be interesting to compare this cycle of events, which is characteristic of nerve and muscle impulses, with the sequence of permeability changes accompanying the rise and fall of the "end-plate potential" at the nerve-muscle junction.

The result of electric excitation is that a small potential change, e.g. a depolarization of 15 mV in the case of a squid axon, becomes amplified several times. This occurs at the expense of the ionic concentration gradients which, though only imperceptibly reduced during a single impulse, would fall slowly during a continued high-rate of activity, but are maintained, *in vivo*, by continued metabolic activity of the cell.[22] The electric amplification is essential, for without it the propagation of nerve signals would fail. This is, indeed, the basis of the action of some local anaesthetics which block nerve impulses by interfering with the electric amplification process of the axon membrane. If a *subthreshold* current pulse (or an inward pulse which displaces the membrane potential in the wrong direction, away from its "ignition point") is passed through the membrane at one point, it produces a local potential difference between inside and outside. This acts as a weak and ineffective signal which rapidly fades out as it spreads along the fibre. The decremental transmission of such subthreshold changes is known as "electrotonic spread" and is analogous to the propagation of electric pulses in a highly leaky and capacitative submarine cable.[28], [29], [31–35] It depends on the cable-like properties of the resting fibre, with its high-resistant membrane and relatively well-conducting "core." By this mode of defective cable-transmission signals can spread over distances of one or a few millimetres, but even within this short range, severe attenuation and "blurring" of the pulse shape occur. But once the signal exceeds, at any point, the threshold of the excitatory relay mechanism, it is locally re-amplified and this process then repeats itself from point to point, ensuring continuous propagation without any loss of signal strength. The cable-mechanism of the nerve fibre, though by itself quite insufficient, forms an integral part in this mechanism of transmission, for it enables an action potential which has arisen at one point of the axon, to repeat itself at other points by imparting to them a sufficient threshold depolarization.[36], [37] This process, which is called "local circuit" transmission, depends on structural continuity of the axon cable and does not provide a means of effective interaction between adjacent cells. Although there are special conditions under which electric currents produced by one cell can directly stimulate another cell,[38], [39] there is no evidence, nor any *a priori* reason to assume, that this happens at natural synapses. In the case of the nerve-muscle junction, experiments will be discussed below which support the opposite view.

VOL.
6
1956

Fig. 2. Photomicrograph (from NASTUK[83]) showing the edge of a frog's nerve-muscle preparation, with a micro-electrode (seen out of focus, as a vertical shadow) about to enter the synaptic region of a muscle fibre. Outlines of muscle fibres can be seen with nerve branches terminating in some of them.

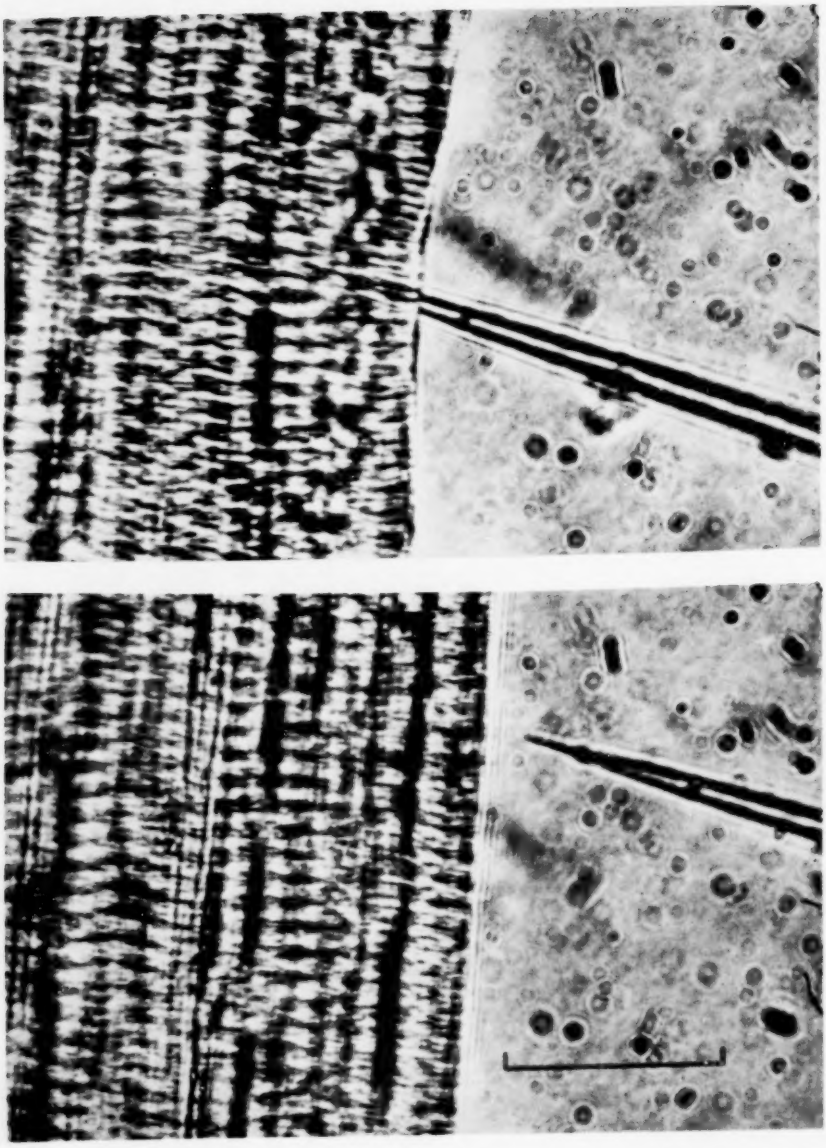


Fig. 3. Photomicrographs of a crustacean muscle fibre, a micro-electrode being inserted between the two exposures. Scale 50 μ. (From FATT and KATZ[116]).

VOL
6
195

B. Chemical transmission from motor nerve endings to muscle end-plate

It was known for many years that the motor end-plate, i.e. the specialized region of the muscle fibre surface which is contacted by the nerve, is highly sensitive to various drugs which can produce local stimulation or paralysis without directly affecting the resting properties or the active response of the nerve and muscle fibres on either side.[40], [41] Blockage by curare, and excitation by nicotine are typical examples. It was later found that one of the most potent chemical stimulants,

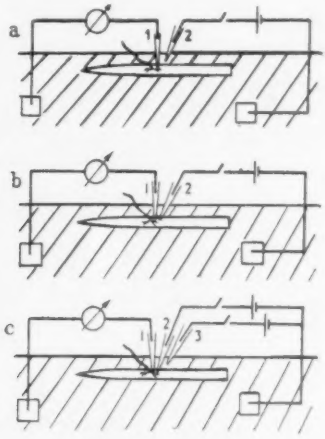


Fig. 4. Diagram of a nerve-muscle junction, showing different applications of the micro-electrode technique (from CASTILLO and KATZ[64]). In all three cases, the membrane potential is recorded between an electrode inside the muscle fibre (1) and the external fluid. Additional micro-electrodes can be used, e.g. to displace the membrane potential by passing current across the fibre surface (*b*), or to apply ionized pharmaco-dynamic substances like acetylcholine$^+$ to the end-plate, either from the outside (*a*) or the inside of the fibre (*b*). Combination of these methods is shown in (*c*). For details of these techniques, see [5], [6], [7], [8], [64], [77].

acetylcholine, is continually synthetized and present in "bound" form in the motor nerve,[42–46] and that it is released from the terminal portions of the nerve during excitation.[47], [48] It was also discovered that a powerful and specific enzyme is present in high concentration at the nerve-muscle junction capable of hydrolyzing acetylcholine and quickly destroying its effect.[49] The localization of the enzyme has been demonstrated by histochemical means,[50] and its local influence on the transmission of the impulse has been elucidated by the use of specific cholinesterase inhibitors. The effect of such inhibitors is to make the excitatory events at the motor end-plate which follow the arrival of a nerve impulse persist for a longer time:[48], [51], [52] this

again is a strictly localized effect and does not occur at other points of the nerve-muscle system.

Evidence of this kind has led to the view that the transmission of an impulse from nerve to muscle is mediated by a chemical stimulant, acetylcholine, which is released from the pre-synaptic terminals and acts by depolarizing the post-synaptic membrane. Within the boundaries of a single cell, impulses spread by electric currents from one region to the next, the link between adjacent points being provided by the cable-structure of the fibre. According to the chemical transmission theory,

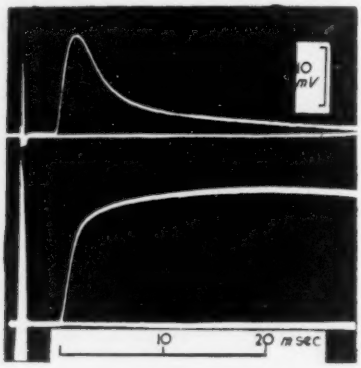


Fig. 5. The prolongation of the neuro-muscular transmitter action by a cholinesterase inhibitor (from FATT and KATZ[7]). The records are local end-plate responses to single motor nerve impulses and were obtained by intracellular recording. By reducing the Na-content of the bath to one-fifth, the transmitter effect has been reduced in intensity below the level at which muscle spikes are initiated, and simple localized end-plate potentials are therefore recorded (cf. Figs. 15 and 17). The lower record shows the striking change in the time course of this response under the influence of an anticholinesterase (prostigmine, 10^{-6} weight/vol.).

this electric link is broken at the synapse, and the local circuit action of the axon has been replaced here by short-range diffusion of a specific chemical substance.

III. Criticisms of the Chemical Transmission Theory

The evidence for a specific transmitter function of acetylcholine at the myoneural junction has been widely but not universally accepted, and attempts have been made[18], [19] to establish a unifying concept according to which acetylcholine plays an essential part in the electric excitation of the axon, and the transmission across the synapse does not differ fundamentally from the propagation of the impulse along the membrane of a single cell. Although the experimental results outlined in the previous two sections appear to render such a view untenable,

it is nevertheless true that there are important gaps in our present information. Until the physico-chemical mechanism of the permeability changes in the axon membrane has been elucidated, the participation in this process of substances like acetylcholine or cholinesterase may be (and has been) postulated without any means of immediate proof or refutation. On the other hand, the evidence for an essential role of acetylcholine at the nerve-muscle junction does not, by itself, rule out the possibility that a direct spread of action currents may take place and play a part in the transmission of the impulse across the synapse. The fact is that the original experiments which established, on one side, electrical transmission in the axon and, on the other side, chemical transmission at the synapse arose along different lines of approach and were designed by their authors to prove one, but not to eliminate the other possibility.

The "unitary" view which has been advocated by NACHMANSOHN and others,[18], [19], [53] requires (*a*) that local circuit action should operate continuously, from nerve to muscle and (*b*) that the release and action of acetylcholine is an "intra-cellular," or "intra-membraneous" event at each point, in the nerve as well as the muscle fibre. It is maintained that there is nothing specific in the synaptic situation, and that findings of a special chemical sensitivity, of local release of acetylcholine, of localized effects of anti-cholinesterases, etc., might all be explained by spatial differences in accessibility and the presence, or absence, at different points of the system, of impenetrable diffusion barriers. It has been claimed that the drugs which have long been known to exert strong effects on the motor end-plate, can be shown, by intracellular application, to be equally effective on the axon membrane.[19], [54] It has further been objected that there is a serious quantitative discrepancy between the very small amounts of acetylcholine which are released from active nerve endings into the perfusion fluid and the relatively large amounts which are required to stimulate the motor end-plates by direct application. Finally, it has been argued that an action which depended on diffusion across the synaptic "gap" would be inefficient and too slow to account for the rapid "non-stop" transmission of the motor impulse from nerve to muscle.[53] It will be appropriate to examine the validity of these claims and objections.

A. Does electrotonic transmission take place across the neuro-muscular junction?

A positive answer to this question has often been taken for granted, and it would indeed be an invidious task to recite all the "electrical" theories in which, tacitly or otherwise, trans-synaptic cable-spread of action currents was assumed, without experimental evidence.

It is true that artificial situations can be created where one nerve

fibre will stimulate, by its action current, another adjacent fibre.[38], [39], [55], [56] But it is equally certain that this does not normally happen, and even when two large non-medullated axons are placed side by side, it is necessary to transfer them from their normal, conducting, fluid environment into an electrically insulating medium in order to make any electrical interaction between them detectable. There is, in any case, little point in arguing about synaptic transmission by analogies from morphologically and electrically quite different situations. The only way to deal with the problem is to try and obtain direct information by experiments on the synapse itself.

The experimental question is whether *subthreshold* electrical signals, or more generally speaking, electric potential changes which do not involve nerve impulses, are transmitted across the neuro-muscular junction. In the nerve axon a subthreshold p.d. produced at one point of the membrane by the passage, e.g., of a constant current, spreads along the fibre and, though greatly attenuated, can be detected a few millimetres away. Unlike the nerve impulse, but like an ordinary cable signal, these attenuated potentials vary in amplitude and electrical sign with the strength and direction of the applied current. Similar experiments were made at the nerve-muscle junction by passing electric currents through the terminal portion of a motor nerve fibre with one electrode close to the synapse, and a recording probe in the end-plate region of the muscle fibre.[57] The result was quite definite: no electrotonic potentials were transmitted across the cell junction. When the applied current exceeded the threshold of the nerve fibre, an action potential appeared at the motor end-plate after a delay of about 1 msec., and even if the current was made many times stronger, the delay could not be reduced much below 1 msec. It appears from these experiments that the direct electrotonic transmission of potential changes comes to a stop at the nerve terminal, and some other process must intervene involving a definite synaptic delay.

There is, however, one criticism which can be directed against this interpretation: the electrode was placed on the nerve close to the cell junction, but there always remained an inaccessible stretch of terminal axon filaments between the point of electrode contact and the myoneural synapse. It could be argued that the attenuation of subthreshold electric signals in these fine filaments is much more severe than in larger and more easily accessible parts of the system, and that therefore the signal might have faded out before reaching the junction. If that were so, the experiment would indeed have failed to tell us anything about the electrical properties of the synapse.

This apparently intractable objection has been disposed of by the following more recent observations: the neuro-muscular junction is in a state of spontaneous local activity which manifests itself in an

intermittent and random appearance of minute electric potential changes ("miniature end-plate potentials") in the post-synaptic membrane.[58] The activity originates in the nerve endings, and pharmacological tests have provided evidence that the phenomenon is due to the intermittent local discharge of acetylcholine from pre-synaptic terminals. It has recently been found that the frequency of this discharge can be enormously increased by passing a steady electric current through the terminal part of the motor nerve.[59], [60] This observation is highly relevant to our present discussion, for it shows that electrotonic spread does proceed along the motor nerve terminals and produces profound alterations at the synaptic contact points of the nerve where the miniature e.p.p.'s originate. The experiment also showed that local-circuit transmission stops at these points, for no trace of electrotonic potentials could be detected on the other side of the nerve-muscle junction, although their effect on the release of acetylcholine and the discharge of miniature e.p.p.'s was obvious.

VOL.
6
1956

B. Are the quantities of acetylcholine released from nerve endings inadequate to stimulate motor end-plates?

The amount of acetylcholine which is released during the transmission of an impulse from the nerve endings is extremely small, and in order to collect measurable quantities in the perfusing fluid, it is necessary to use a cholinesterase inhibitor and to apply repetitive stimulation to the nerve for long periods. Under these conditions, it has been estimated that with each impulse an average amount of one-half to one million molecules are discharged from each myoneural synapse into the circulation.[61] This value probably represents no more than a fraction of the quantity locally released from nerve endings during a single impulse, because losses are likely to occur in the tissue spaces and also as a result of the prolonged period of stimulation. To stimulate the muscle by *applying* acetylcholine, much larger doses are needed, and this has been used as an argument against the chemical transmission theory. What has apparently been overlooked in this argument is that the only relevant quantity is the *concentration* of acetylcholine built up by an impulse at the receptor site, and this quantity is as yet undetermined. Whether a given *amount* suddenly injected into the system produces an effective concentration at the motor end-plate, must depend on the distance over which the substance has to diffuse. For example, if a quantity of acetylcholine is released instantaneously at point A into a large fluid volume, the maximum concentration which is attained at point B is inversely proportional to the third power of the distance AB. As recent electron micrographs have shown,[62] the distance between pre- and post-synaptic structures at the myoneural junction is less than one micron, and at many points there appears to

be intimate contact between the membranes of the nerve terminal and the muscle "end-plate."

It is not practicable to create an identical situation by means of an external applicator; nevertheless it is possible to eject acetylcholine electro-phoretically from a pipette of ultra-microscopic tip size, and in this way small controlled quantities can be applied to a motor end-plate at very close range and in a very brief interval of time.[8], [63], [64] With the help of this technique, artificial "end-plate potentials" (or "acetylcholine-potentials") have been produced which differ but little from those set up via a nerve impulse except in showing a much slower time course. The rising phase of the quickest ACh-potentials so far obtained was about 15 msec., which indicates that the length of the diffusion path for the applied ACh cannot in practice be reduced much below 10 microns, or that the receptors are surrounded by a zone of slow diffusion.[65]*

Under these nearly optimal experimental conditions, a dose of ACh of the order of 10^8 to 10^9 molecules was sufficient to set up a muscle impulse. This is some hundred times greater than the quantity released from the nerve endings, a difference which is not surprising in view of the different pathways and distances to be overcome by diffusion. There are, at present, too many uncertainties attached to these calculations to pursue them in any detail; but they make apparent the irrelevance to the transmitter argument of the alleged discrepancy between applied and released amounts of acetylcholine.

To summarize the position of the problem, we find (*a*) that there is no experimental support for the claim that electric transmission can proceed across the neuro-muscular junction. On the contrary, there is good evidence that electrotonic currents fail to spread across this synapse. (*b*) The claim that not enough acetylcholine is released from the nerve endings to stimulate the motor end-plate appears to have no valid foundation and to have arisen from a neglect of the relevant quantitative factors. The further assertion which has often been made, namely that diffusion across the synapse is too slow to account for the high speed of neuro-muscular transmission seems to rest on a similar misconception. Short-range diffusion is a rapid process and, with a separation between source and receptor of less than a micron, the concentration of acetylcholine at the receptor site will build up to a peak within a few hundred microseconds after the release of the transmitter. The time taken up by diffusion of this substance across the synapse may account for only a small portion of the total synaptic delay and ascending time of the end-plate potential.

* For the simplified case of instantaneous point source and point receptor separated by distance r, the time T required for the concentration to rise to a peak is $r^2/6k$, where k is a diffusion constant, for ACh about 8×10^{-6} cm^2 sec^{-1} in free solution.

VOL
6
195

C. On the effect of intra-cellular application of acetylcholine and related substances

The work outlined on p. 127 has shown that the motor end-plate possesses a high and specific sensitivity to various pharmaco-dynamic agents (see also [66], [67]). In its reactions to a blocking substance like curarine or to an anti-cholinesterase the synapse differs from other parts of the muscle fibre qualitatively, not just in the order of magnitude of the required dose. It has nevertheless been argued that such a difference might be explained by the presence of diffusion barriers around the fibre surface which prevent access of the drug to a reactive area of the membrane, but which may be absent at the neuro-muscular junction. In support of this argument, evidence was presented to show that cholinesterase-inhibitors are able to block the propagation of impulses along nerve and muscle fibres, provided their chemical constitution enables them to penetrate the surface membrane and to build up a high enough concentration inside the fibre to reduce the activity of the intra-cellular enzyme.[18], [19] The external concentration of the inhibitor in these experiments exceeded that used "conventionally," in the earlier work on the synapse, by several orders of magnitude, but this was regarded as irrelevant. The significance of these findings may, however, be questioned on several other grounds: in the first place, the effect of the anti-esterase on the axon differs qualitatively from the effect on the synapse where the principal result is a striking prolongation, and not a block, of the local transmitter activity.[48], [51], [52], [68] This is an important difference, for prolongation and not blockage is the obvious consequence predicted by the theory. Furthermore, if the muscle is stimulated directly, not through its nerve, the impulse passes the synaptic region of the muscle fibre without any change in its properties, even when transmission from nerve to muscle has been strikingly modified by the esterase inhibitor.[48], [68a] These observations favour the concept of a specific role of acetylcholine in synaptic transmission.[57], [68] The results quoted by NACHMANSOHN[19] permit of various interpretations: it may be that intracellular metabolism of acetylcholine is connected with the maintenance of excitability, or perhaps the excessive extra-cellular concentration of the drug had some unspecified effect on the surface membrane.

More recently, however, NACHMANSOHN[19] reported the interesting observation that the propagation of an impulse along the nerve axon may be stopped by the *intracellular* application of a minute quantity of curarine. This seemed to provide direct and conclusive support for the idea that acetylcholine or some related substance is intimately associated with the excitatory process in the axon membrane, and that previous failures to demonstrate it were due to inadequacy of the

conventional methods of drug application. It is difficult, however, to assess the evidence because it has not been published except in a preliminary form (GRUNDFEST *et al.*[54]), and recent attempts by HODGKIN and KEYNES[69] to confirm the findings reported by GRUNDFEST *et al.* have given negative results.*

At the motor end-plate, direct evidence has recently been obtained of precisely the opposite situation, namely that the receptor molecules

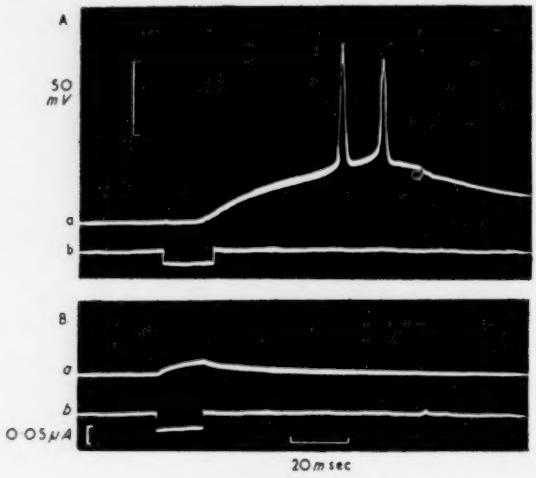


Fig. 6. External and intra-cellular application of acetylcholine to a motor end-plate (from CASTILLO and KATZ[117]). In *A*, an ACh-filled micropipette was placed on the outside of an end-plate, and a quantity of ACh was released by passing a brief outward-directed current pulse through the pipette (registered in trace *b*). It produced the effect shown in trace *a*: a depolarization, developing after a diffusion delay and culminating in two spikes. Between records *A* and *B*, the ACh-pipette entered the muscle fibre. An outward pulse produces now a small catelectrotonic potential, due to the passage of electric current across the fibre membrane. (If, for comparison, a KCl-filled micropipette is used, no potential change is produced by the pulse until the pipette has entered the fibre, when the effect is identical with the electrotonic potential recorded in *B*, *a*).

with which acetylcholine reacts, are situated on the external surface of the cell membrane and are—at least relatively—inaccessible from the interior of the fibre.[63], [64] Electrophoretic ejection of acetylcholine from a micropipette is very effective when the tip of the pipette is placed on the outer surface of the end-plate (that is the surface which faces the nerve endings). But as soon as the pipette enters the muscle fibre, application of the same dose of acetylcholine becomes ineffective. It might be supposed that this difference could be due to a higher

* In an addendum to a recent paper[69a] GRUNDFEST himself states that he has failed to confirm the earlier observations with curarine.

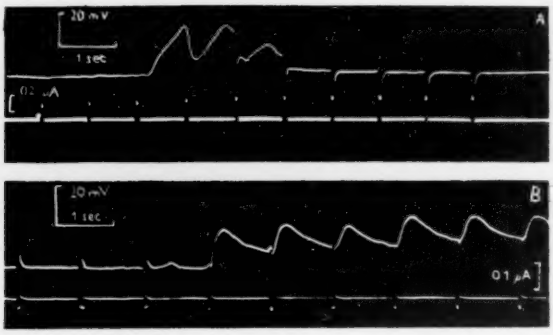


Fig. 7.

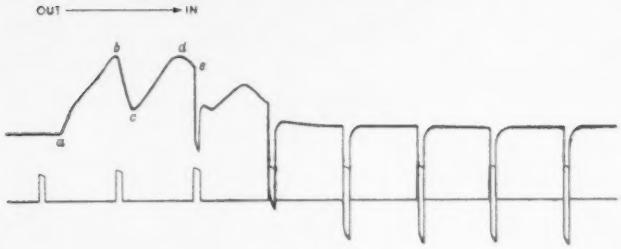


Fig. 7*a*.

Fig. 7. The effect of inserting and withdrawing an ACh-filled micropipette at a motor end-plate. (From CASTILLO and KATZ,[64] where the procedure is explained in detail.) *A* (redrawn for clarity in Fig. 7*a*): the pipette approaches the end-plate from the external bath. At regular intervals, electric "braking" pulses are passed into the pipette stopping, temporarily, outward diffusion of ACh^+ ions. The current pulses are registered in the lower part of the record, the membrane potential change in the upper part (depolarization is shown as upward deflection). At *a* (Fig. 7*a*), a depolarization develops due to diffusion of ACh from the approaching tip. At *b*, a "braking pulse" is applied and temporarily stops the outflow of ACh. The membrane potential recovers for a moment, but soon efflux of ACh recommences and the depolarization builds up again (at *c*). During this interval the pipette advances into the fibre, shown by the fact that the next pulse produces an electrotonic potential (at *e*). During the next two pulses, the electrotonic potentials increase in size to a steady level, indicating that the tip becomes more firmly "sealed" in the fibre membrane. At the same time, the ACh-induced depolarization subsides. The moment at which the tip first penetrated the fibre membrane is suggested by the decline of the depolarization at point *d*, just before the first electrotonic potential appeared.

B: From another experiment. The ACh-pipette had been inserted into the end-plate region of a muscle fibre, and outward pulses were passed through the pipette producing ordinary catelectrotonic potentials. The record shows the changes which accompanied the withdrawal of the pipette from the fibre interior. This is signalled by the disappearance, or sudden reduction, of the brief catelectrotonic potentials (between the third and fourth pulse) and the simultaneous reappearance of the delayed ACh-potential. For further details see [64]

concentration of cholinesterase inside the cell, but against this is the fact that the same observations were made when the pipette contained a stable substance, carbaminoylcholine, which acts like acetylcholine but is resistant to cholinesterase.

Thus the balance of the evidence is overwhelmingly in favour of the original theory of chemical transmission at the myoneural junction:

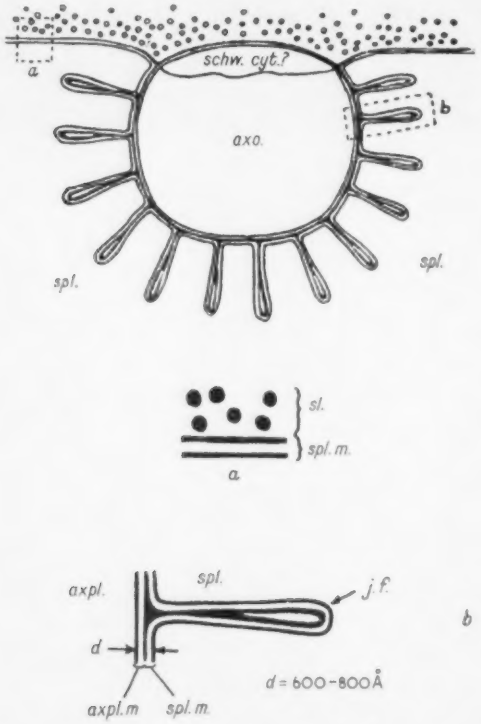


Fig. 8*a*. (see legend to Fig. 8).

schw. cyt.—Schwann cytoplasm; *axo.*—axoplasm; *spl.*—sarcoplasm; *sl.*—sarcolemma; *spl. m.* sarcoplasmatic membrane; *axpl.*—axoplasm. *axpl. m.*—axoplasmatic membrane; *j.f.*—junctional fold.

local circuit transmission stops at the nerve terminals and causes the release of acetylcholine. By short-range diffusion, over distances of less than 1 μ, ACh reaches and reacts with specific receptors located on the outer surface of the motor end-plate.

The problems which present themselves at this stage are the nature of the mechanisms by which acetylcholine is released pre-synaptically and by which it changes the properties of the post-synaptic membrane. This is a virtually unexplored field, but investigations during the last few years have begun to shed some light on certain aspects of these processes.

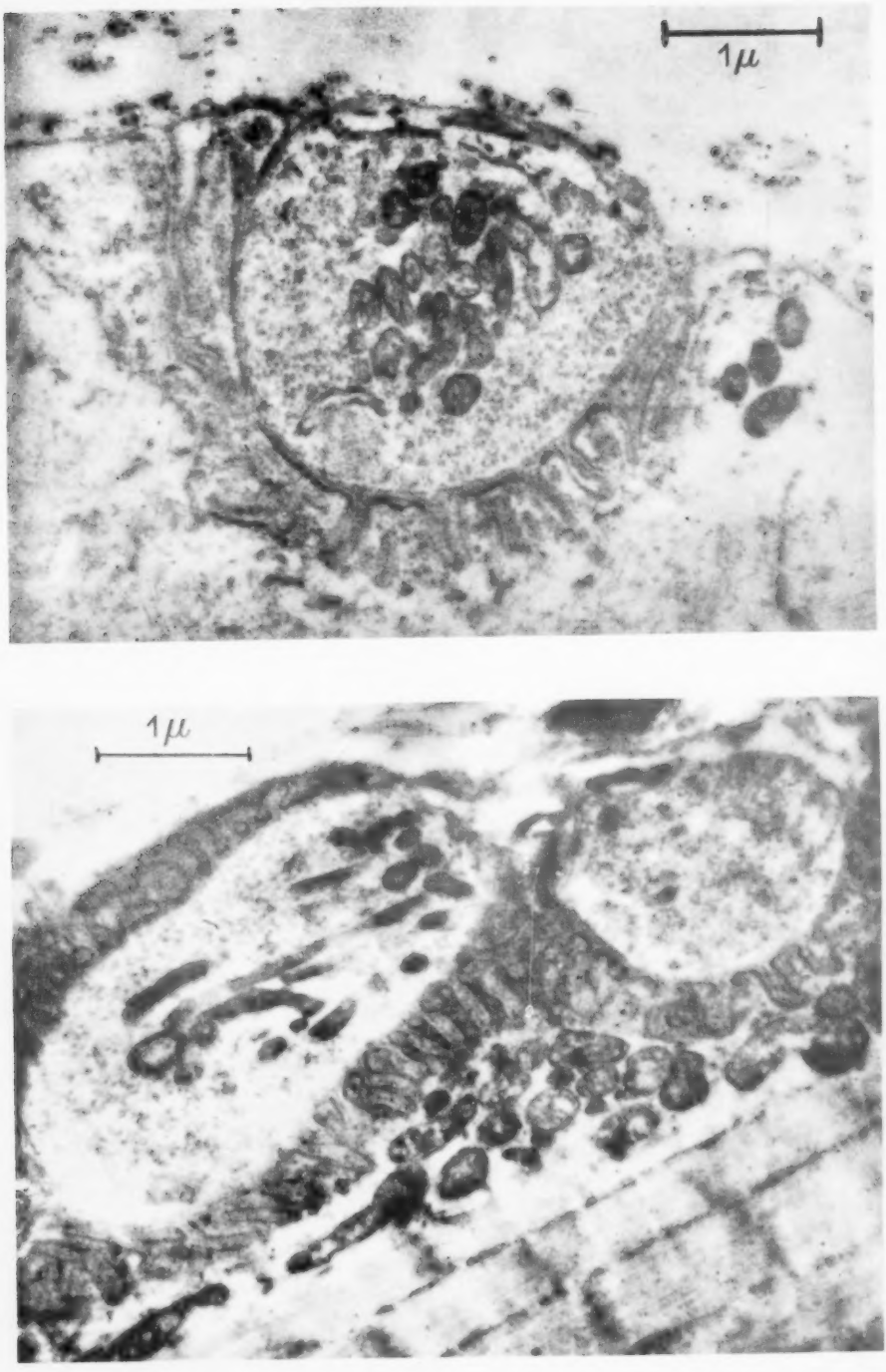


Fig. 8 Electron micrographs of a reptilian neuro-muscular junction (see ROBERTSON[68a]). The material for this figure and Fig. 8a was kindly provided by Dr. ROBERTSON at a time when his paper was being prepared for the press. The axoplasm of the nerve endings contains, in addition to mitochondria (large dark particles), many small globules of less than 0·1 μ diameter.

VOL
6
19

IV. The Different Stages of Neuro-muscular Transmission

A. The release of acetylcholine from motor nerve endings

The general chain of events and the number of steps which intervene between the arrival of a motor nerve impulse and the contraction of the skeletal muscle fibre is represented by the following scheme: Nerve impulse $\xrightarrow{1}$ acetylcholine $\xrightarrow{2}$ end-plate potential $\xrightarrow{3}$ muscle impulse $\xrightarrow{4}$ contraction.

Although a good deal is known about each of these five links, it would be wrong to suppose that the main pieces in the myoneural "jig-saw puzzle" have now been assembled and only minor detail remains to be filled in. The transition (3) from e.p.p. to muscle impulse, that is from the local post-synaptic depolarization to the propagated action potential of the fibre, is well understood.[70], [71] A certain amount has also been learnt recently about stage 2 (the production of the end-plate potential by acetylcholine). But the mechanisms of stages 1 and 4, that is the immediate "effector" actions of the nerve and muscle impulses, have remained as mysterious as ever. The process by which ACh is liberated from the nerve terminals appears to involve several intermediate steps, and there is little doubt that as more information becomes available on the transitional stages, the simple scheme will have to be expanded considerably. The mechanism of ACh-release is difficult to explore mainly because the pre-synaptic structures do not lend themselves to direct experimentation. Inferences about the physical state and chemical properties of nerve endings have had to be made by indirect argument and from analogies, and are therefore likely to be at fault.

One of the important points to consider here is that release of ACh^+ ions differs in an essential respect from the release of, say, K^+ ions which occurs during the falling phase of the axon impulse: ACh, unlike K, is synthetized locally by an intracellular enzyme system and is known to exist, not in free ionic solution, but in a state either of chemical combination with, or structural enclosure in, small particles inside the cell. Recent electron microscope studies[62] have shown that the motor nerve terminals contain a fairly dense population of microsomes, granules or vesicles of less than 0·1 μ diameter, which may well be the intracellular corpuscles to which ACh is attached. It has been known since Loewi's investigations[45], [46] that most of the ACh which is present in "homogenized" nerve tissue can only be extracted into an aqueous solution after chemical destruction of the cell proteins. It appears then that the discharge of ACh from a nerve terminal requires the disruption of more than one diffusion barrier: first the release from its intracellular attachment, and secondly a passage through the nerve membrane.

One might suppose that when a "critical" collision occurs between an intra-cellular "ACh-carrier" and the membrane of the nerve terminal, the two barriers are opened simultaneously and the ACh-contents of the carrier particle are suddenly discharged. An attempt is made to present this schematically in Fig. 9 where an effective collision is distinguished from others by the stipulation that certain labelled spots on the two surfaces should meet. This picture, though purely specu-

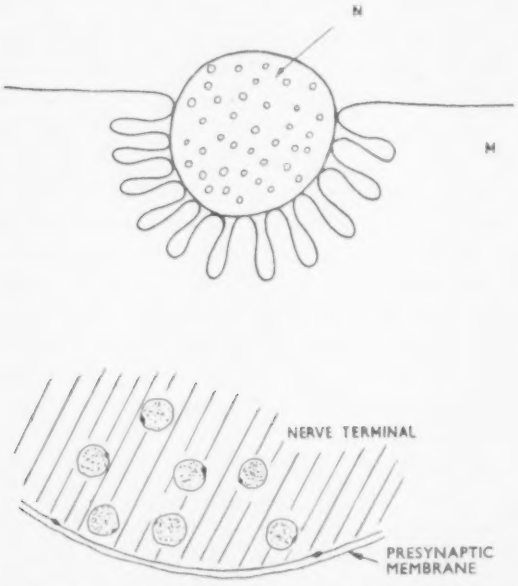


Fig. 9. Diagram of nerve-muscle junction. Several of the features described by ROBERTSON[62] have been included. N, nerve terminal; M, muscle fibre. In the lower part, an enlarged part of the nerve terminal is shown, containing "ACh-carrier corpuscles." Release of ACh is supposed to occur as a result of critical collisions between corpuscles and membrane. This is indicated formally by labelling certain "critical spots" on the surfaces of both. The effect of the nerve impulse might be visualized as temporarily increasing the "critical area" of the pre-synaptic membrane.

lative, is nevertheless in accord with recent experimental findings; it takes account of the evidence discussed below that the release of ACh from nerve terminals occurs in multi-molecular units or "quanta," and of the evidence, already cited, for the bound state of intracellular ACh content. On this scheme there would be no difficulty in understanding the recently reported occurrence of a spontaneous quantal discharge of ACh,[58] but no explanation is offered of the mechanism by which nerve impulses control or facilitate this event. (The immediate influence of the nerve impulse is not likely to disturb the intracellular

constituents, but it might be thought to alter the properties of the membrane in such a way that the density of the reactive spots increases for a moment.)

1. *The role of inorganic ions in the release of acetylcholine*

In an attempt to find the link between nerve impulse and ACh-release mechanism, the role of the inorganic cations which play an important part in conduction along axons has been discussed by several authors.

(*a*) *Potassium.* There is evidence that an excess of potassium ions applied externally causes a liberation of ACh from the nerve endings. This was found in experiments on the perfused sympathetic ganglion[72] and it has been suggested that a "mobilization of potassium ions" believed to be associated with the nerve impulse might be responsible for the normal release of the transmitter. According to our present concepts the movement of potassium through the axon membrane is greatly accelerated during the action potential, but it is not clear why this should bring about the release of ACh from a cellular medium whose concentration of potassium ions is higher than it is anywhere else. Furthermore, the release of ACh observed with potassium application does not necessarily mean that these ions play a direct or specific part, for an excess of potassium is known to depolarize the cell membrane and might therefore simply start off the unknown chain of events by which normally a depolarization of the nerve ending leads to a liberation of the transmitter. This point was emphasized by Brown and McIntosh[73] who found that the effective dose of KCl applied to the ganglion was even sufficient to excite the pre-synaptic axons. More recently, Hutter and Kostial[74] observed that the potassium-induced release of ACh still occurred in a ganglion which had been made inexcitable by perfusion with a Na-free sucrose solution. Nevertheless even under these conditions, potassium depolarizes the cell membrane, and the release of the transmitter which follows may therefore be an indirect effect of K ions.

(*b*) *Sodium.* The suggestion has been made that the efflux of ACh from the nerve terminals may be analogous to that of potassium ions from other parts of the system, and that ACh^+ may leave the cell by specific exchange for sodium ions which enter the terminal during its electric activity.[75] There is, however, little evidence in favour of this hypothesis. One would expect the release of ACh to be diminished when the external Na-concentration is lowered. It is true that the end-plate potential can be reduced and myoneural transmission stopped in Na-deficient media, but this effect may be largely or wholly due to a post-synaptic blocking action of Na-lack, contrary to the original interpretation by Fatt and Katz[75]; (see Nastuk,[8], [76] Castillo and Katz[77]). Direct estimates of the ACh-output from stimulated

preganglionic nerve endings reveal no relation to the Na concentration in the perfusion fluid, down to one-third of the normal level.[74] This does not necessarily dispose of the hypothesis because it is not certain whether the total quantity of ions exchanging across the active nerve membrane is much reduced in the sodium-deficient medium (the amplitude and rate of the action potential is diminished, but its duration is lengthened which may provide a compensating factor). There are other grounds for supposing that indirectly, by altering the properties of the pre-synaptic spike, the sodium concentration would influence the output of ACh from the nerve endings (cf. the anelectrotonic effect observed by CASTILLO and KATZ[60]), but the hypothesis of a direct and specific exchange between Na^+ and ACh^+ has no support. Perhaps the strongest evidence against this concept comes from the recent observation[77], [78] that the spontaneous "quantal" mechanism of the ACh-discharge continues to operate in the absence of NaCl.

(c) *Calcium and magnesium.* While the search for a specific effect of Na and K has not been very fruitful, there is convincing evidence that calcium and magnesium ions exert a direct and specific action on ACh-release. It had been known for many years that Mg blocks neuro-muscular transmission. Only recently, however, has it been established that this blockage is due principally to an interference with the pre-synaptic release of the transmitter substance.[79] This was found by analysis of electrical events at the myoneural junction and has been confirmed by the more direct method of assaying ACh-output from the perfused sympathetic ganglion.[80] It is of special interest that Ca opposes the effect of Mg and relieves the block. Withdrawal of Ca, on the other hand, acts like addition of Mg and produces junctional blockage, through a lowering of the ACh-output.[58], [80], [81] Over a wide range of concentrations, a balance between the two effects can be obtained suggesting a direct "competitive" antagonism between Ca and Mg ions at some critical phase of the pre-synaptic events. There are indications that at other types of nerve endings at which a different transmitter is released, Ca and Mg may have mutually antagonistic effects of the same kind, and that this antagonism may be a rather general synaptic phenomenon. It is not shared, however, by the properties of the nerve axon or muscle fibre whose excitabilities are changed in the same direction by both ions.

One of the most striking effects of Mg, and of Ca-lack, is that the transmitter output is reduced in discrete steps.[58], [82] At a sufficiently low level, the end-plate potential (i.e. the post-synaptically recorded effect of the transmitter) displays quantal fluctuations revealing the statistical composition of the response at each junction. It is significant that the ultimate units which can be recorded at a single end-plate are identical in amplitude and shape with the spontaneously occurring

VOL
6
195

miniature e.p.p.'s which are discharged at random intervals, in the absence of nerve impulses.

2. *The spontaneous quantal release of acetylcholine*

It has become clear during recent years that the "resting" myoneural junction is the seat of spontaneous intermittent activity on a small scale, resulting in the appearance of minute e.p.p.'s at irregular intervals.[58] As this form of electric activity remains well below the threshold of the fibre membrane, it remained undetected until modern

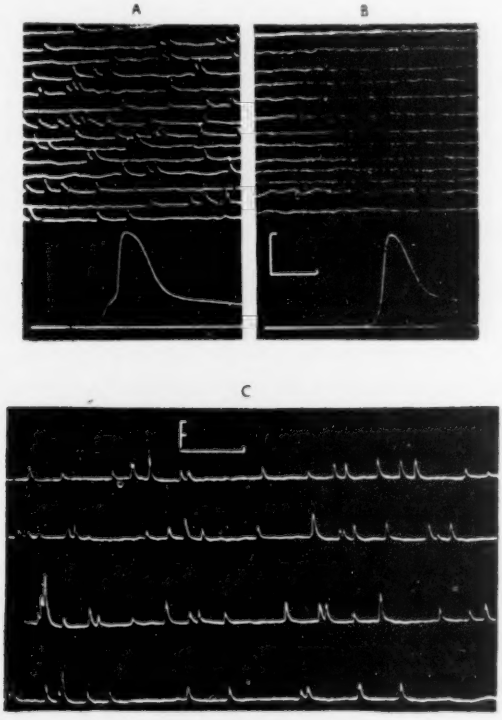


Fig. 10. Spontaneous miniature end-plate potentials at the myoneural junction. (From FATT and KATZ.[58]) The records were obtained with intracellular recording. *A*: Micro-electrode was placed inside a frog muscle fibre at the nerve-muscle junction. *B*: Electrode was placed 2 mm away into the same muscle fibre. The uppermost portions were recorded with slow speed and high amplification (calibrations: 3·6 mV and 47 msec.); the lower portions of *A* and *B* show the response to a nerve impulse with fast speed and low-gain recording (calibrations: 50 mV and 2 msec.). The stimulus was applied to the nerve at the beginning of the sweep; response *A* (at the end-plate) shows the step-like initial e.p.p., response *B* shows the propagated action potential wave, delayed by conduction over a distance of 2 mm. The figure shows that the spontaneous activity is restricted to the myoneural junction.

Lower record: A series of miniature potentials from a prostigmine-treated end-plate. Voltage and time scales: 2 × 1 mV and 1 sec.

refinements in recording technique became available. There has been much evidence to show that these "miniature e.p.p.'s" are due to a localized bombardment of the post-synaptic membrane by packets of ACh released spontaneously from the nerve terminals.

The amplitude of the miniature e.p.p. can be altered experimentally over a wide range: it can be increased by anti-cholinesterases and reduced by curare-like substances.[58] The various agents which have been found to influence the size of the spontaneous potential are all known to affect the sensitivity either of the receptors or of the cholinesterase, or to interfere in some way with the electrical properties of the muscle membrane. None of the observed changes in the amplitude of the miniature e.p.p. seems to reflect a change in the pre-synaptic output of ACh: the quantity of ACh which is released during each spontaneous event appears to be relatively constant. (This is, of course, a tentative generalization, but it probably applies even to the effects of sodium withdrawal and of procaine[58] which, though at one time believed to be associated with a diminished output of ACh, have also been shown to produce large desensitizations of the end-plate receptors ([76], [77], [83] and unpublished observations by CASTILLO and KATZ).)

How are we to interpret this "unit of ACh-release," and what is the mechanism of its spontaneous discharge? The problem would be simple if each unit represented the discharge of a single molecule of ACh. This would immediately explain the constancy of the "quantum" under a variety of conditions and reduce the mechanism of the spontaneous discharge to a random leakage of ACh molecules from the nerve endings.

However, this view cannot be maintained in the light of the following evidence:

(*a*) If a single molecule of ACh were to produce a discrete miniature e.p.p. of standard size and time course, then the application of an ACh-solution to an end-plate should enormously increase the frequency of the random discharge. This is not the case; the effect of applied ACh is to produce a relatively slow depolarization whose amplitude and time course are continuously graded, depending upon the dose of ACh and the time required for diffusion.[58], [64], [84] The molecular components of this effect cannot be detected, presumably because they are far too small and too frequent to be resolved by our recording instruments. By contrast, the discrete size and rapid time course of the miniature e.p.p. indicates that it must be due to a brief synchronous action of many ACh molecules, and requires the sudden release of a whole "parcel," at very short range. The validity of this comparison might be questioned, for it could be argued that, unlike the naturally released transmitter substance, applied ACh may not reach the synaptic receptors and may act on other parts of the membrane with different

characteristics of response. But this argument is contradicted by the fact that competitive blocking molecules, like the quaternary bases of curare, can undoubtedly reach the synaptic receptors and combine with them although their molecular size exceeds that of ACh.

(*b*) If each miniature e.p.p. were to arise from the transient reaction with a receptor of a *single* ACh molecule, it would be difficult to understand the graded effects of inhibitors, and anti-cholinesterases, on the size of the potentials. One would expect a "competitive inhibitor" like curarine to interfere with the reaction in a statistical, all-or-none, manner, that is to reduce the probability, and the recurrence rate, of the spontaneous potentials without affecting their amplitudes. Precisely the opposite effect is observed: curarine reduces the size of the miniature e.p.p.'s, but not their frequency.[58]

On a similar argument, an anti-cholinesterase, by increasing the average life-time of released ACh molecules, would be expected to increase the chances and therefore the frequency, of the discharge without altering the individual potentials. An increase of frequency was sometimes observed,[58], [85] but this is an irregular phenomenon and of uncertain significance. The main and consistent action of an anti-esterase is to increase the amplitude and duration of each miniature potential.[58] The miniature e.p.p.'s have a uniform time course with usually little variation between them. Presumably, transmitter molecules continue to react with end-plate receptors for the duration, at least, of the rising phase of the miniature e.p.p. It would be difficult to conceive of a single molecule remaining combined with the receptor for approximately the same length of time in every instance, without large random fluctuations of this interyal.

(*c*) The quantity of ACh which is released by one impulse at a single nerve-muscle junction has been estimated as of the order of 10^6 molecules[61] or more.[11] This produces a full-size e.p.p. which is about 100 times larger than the miniature e.p.p. Although it would be wrong to assume proportionality over any wide range between the quantity of released ACh and the size of the e.p.p., nevertheless the comparison indicates that a great many molecules of ACh are needed to build up a miniature potential.

We are thus forced to conclude that the release of ACh, responsible for the miniature potential, occurs in large quanta, each containing hundreds or thousands of molecules, and the problem arises how such an event can take place spontaneously. This question has given rise to a good deal of speculation and also to several new experiments.

The suggestion has been put forward that the miniature e.p.p. may be the outcome of a localized spontaneous excitation in the nerve endings.[58] This seemed plausible because spontaneous fluctuations of the membrane potential ("thermal agitation noise") would be largest

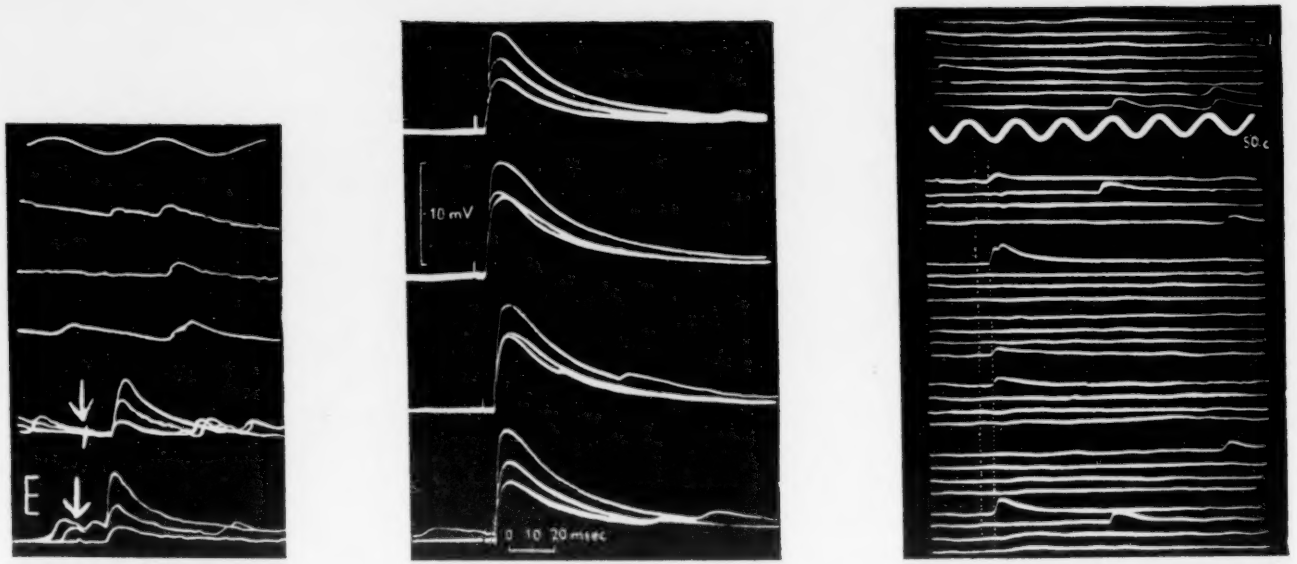
in small nerve endings and might momentarily exceed the threshold of excitation at random points, without necessarily giving rise to a propagated impulse in the axon. On quantitative grounds, however, it was doubtful whether this reasoning could be applied to the motor nerve terminals.[84] The suggestion has been finally disposed of by the recent discovery that the spontaneous discharge at the myoneural junction continues even when the tissue has been depolarized and made inexcitable by immersion in an isotonic solution of K_2SO_4.[77], [78]

An alternative hypothesis was that the release of ACh from the nerve ending may depend on the activation of "key" or "carrier" molecules in the pre-synaptic membrane.[60], [84], [86] These may be likened to localized gates which open and shut in an all-or-none manner and instantly release a flood of ACh ions from the interior of the cell. Random activation of single "carriers" would occur spontaneously; the nerve impulse lowers their activation threshold and thereby synchronizes the activity of a large number of ACh-carriers.

The idea of a specific ACh-carrier in the pre-synaptic cell membrane was derived from similar hypotheses concerning the "activated" flux of Na and K ions across the axon membrane during the passage of an impulse.[2] However, another and perhaps much simpler explanation for the quantal nature of ACh-release is suggested by the presence of submicroscopic granules or vesicles inside the nerve terminal which may be the normal "carriers" of intra-cellular ACh.[62] It is conceivable that such particles might either be discharged, like secretion granules, or alternatively lose the whole of their ACh content when they reach the outer surface of the nerve membrane. It seems pertinent that the "quantum of release" is relatively constant and uninfluenced even by grossly abnormal changes in the properties of the cell membrane (for example during the anelectrotonic "breakdown" effect[60] and the K_2SO_4 treatment[77], [78]); on the other hand the probability—and frequency—of its occurrence, can be varied experimentally over an extremely wide range by several procedures, for instance by altering the membrane potential of the nerve endings. In a general way, it seems that the statistical *chances* of ACh-release are governed by the labile properties of the cell membrane, while the *quantum* of release may owe its stability to the way in which ACh is distributed inside the cell in discrete structural parcels. The standard size of this parcel may be determined by the amount of ACh which can be built up and accumulated within a single microsome concerned with its synthesis (cf. [42–44]).

3. *The relation between "responsive" and "spontaneous" release of ACh*

The spontaneously occurring miniature e.p.p. represents the effect of a single quantum of ACh; the normal e.p.p. response is built up by a large number of such units. Evidence for this view has been obtained



(*a*) (*b*) (*c*)

Fig. 11. *Quantal fluctuations of response at a single nerve-muscle junction.* By raising the magnesium or lowering the calcium concentration, the e.p.p. response to a nerve impulse is reduced to a small number of quantal units. During successive impulses, such units are brought into action in a random manner; hence the large fluctuation in the size of the response. In Fig. 11c, it becomes evident that the fluctuations in the size of the e.p.p. are discontinuous and involve "quanta" of a size identical with that of the spontaneous miniature potentials. 11*a*: (from FATT and KATZ[58]). The upper three sweeps show spontaneous discharges; in each of the lower two records 3 superimposed sweeps are shown, during each of which a nerve impulse was set up at the moment marked by an arrow. The e.p.p. response consisted of one or a few "units." Voltage scale: mV; time signal 50 c/s. In 11*b* (from CASTILLO and KATZ[82]) the reduction of the response was less severe, the mean size of the e.p.p. being many times larger than the spontaneous potentials. Each record contains 3 superimposed sweeps showing e.p.p. responses (at a fixed time of the sweep) and spontaneous activity (at random moments). In 11c (CASTILLO and KATZ[82]) the *mean* number of responding units was much less than unity, there being a high proportion of failures of response. The top part shows a few spontaneous potentials. The lower part (below the 50 c/s time signal) shows responses to single nerve impulses. The moment of nerve stimulation is shown by the first dotted line, the latency of the e.p.p. response by the second line. Only 5, out of 24, nerve impulses produced a response.

by analyzing the statistical fluctuation of the responses at a single junction.[58], [82], [87], [88] Within an appropriately chosen range of experimental conditions, the distribution of amplitudes was found to be discontinuous and to correspond closely to a Poisson-grouping of the contributing units.[82]

In particular, when the mean number m of responding units has been reduced to the order of unity (by withdrawing Ca or adding Mg), the proportion of total failures is very close to the predicted value e^{-m} and the results, in a large number of successive trials, agree with the theoretical equation[82]

$$m = \frac{\text{mean amplitude of e.p.p. response}}{\text{mean amplitude of miniature e.p.p.}}$$
$$= \log_e \frac{\text{number of nerve impulses}}{\text{number of failures of e.p.p. response}} \quad \ldots.(1)$$

It appears then that the pre-synaptic apparatus is composed of a population of units which may discharge individually, at random, or synchronously, after arrival of a motor nerve impulse.

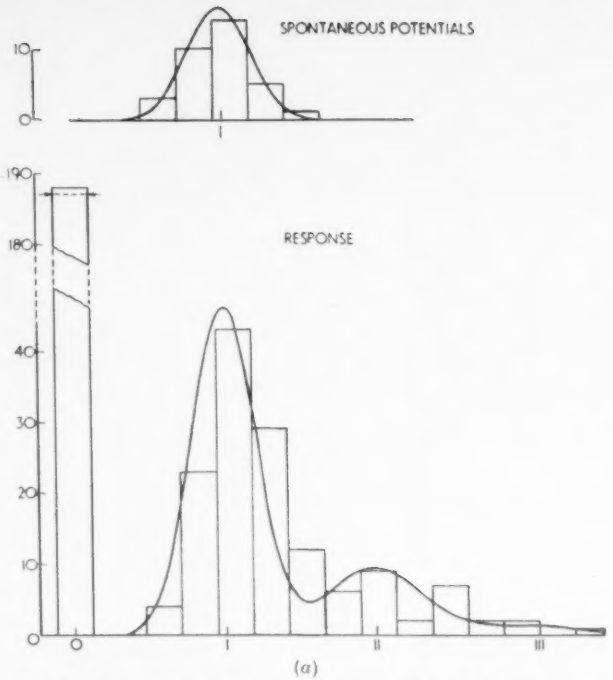
While the statistical relationship between the spontaneous miniature e.p.p. and the full-size response seems clear, the reason why a nerve impulse brings about synchronized activity of many units is obscure. Withdrawal of Ca or addition of Mg reduces the e.p.p. "statistically," by cutting down the size of the responding pre-synaptic population. The size of each quantum of ACh is not affected by either Ca or Mg.[58], [86] It appears that, between the arrival of a nerve impulse and the release of any individual unit of ACh, there is an intermediate reaction involving a calcium compound and unable to proceed if calcium has been displaced by magnesium. This, at least, is the most probable interpretation of the antagonism between these two ions.

The rate of the *spontaneous* activity is, rather unexpectedly, not affected in any consistent manner by calcium or magnesium.[58], [86] This points to an interesting difference between the spontaneous and responsive mechanisms of ACh release for which an adequate explanation has yet to be found (for a discussion of this matter see [60], [86]).

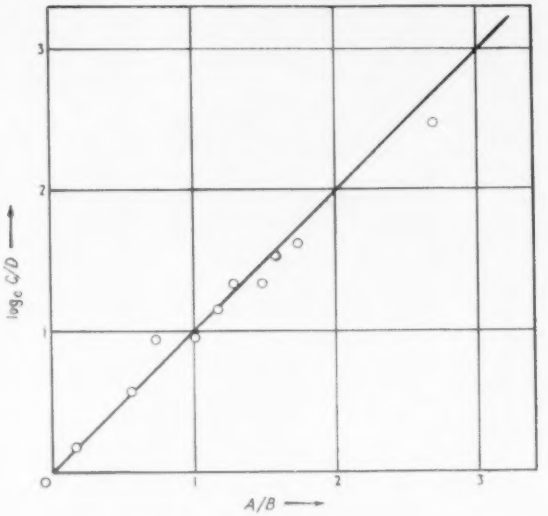
4. *Other factors controlling the release of ACh*

Calcium and magnesium are not the only agents which affect the liberation of the transmitter. Other known factors are the effects of preceding stimulation ("facilitation" and "depression"[87], [88]) and of local electrotonic alterations of the nerve endings.[60] A result of rather general importance has emerged from these studies, namely that all the pre-synaptic effects were associated with a statistical change in the *number* of quantal units. Thus, facilitation (the increase in amplitude

PHYSIOL. 1956



(a)



(b)

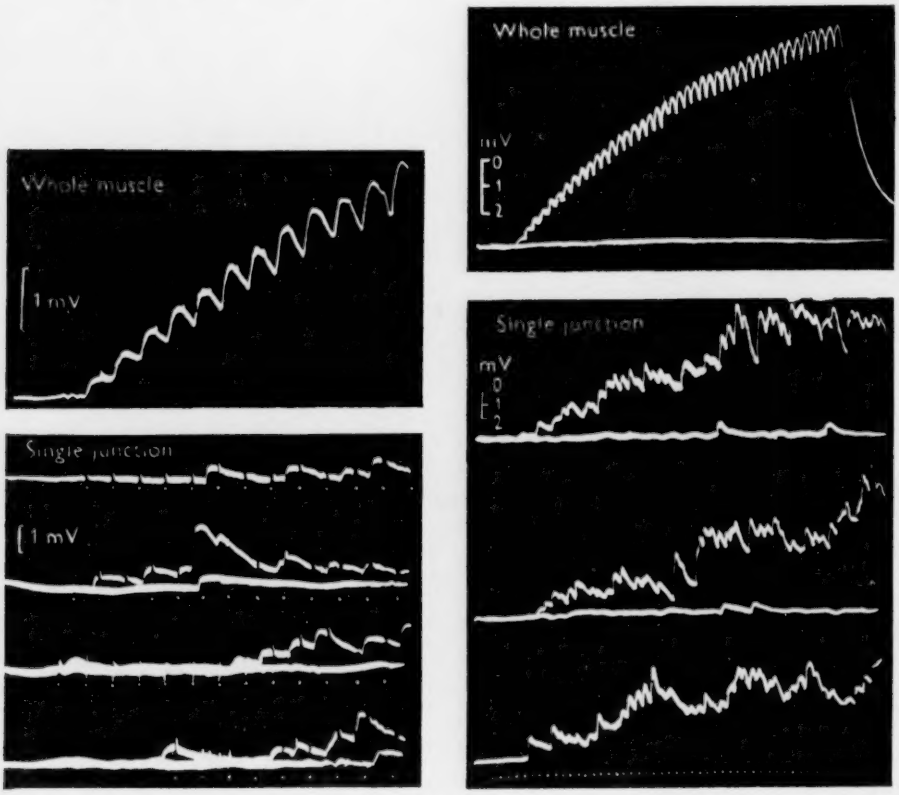
Fig. 12. Statistical analysis of responses at single nerve-muscle junctions. (From CASTILLO and KATZ.[82]) (a) Histograms from a single experiment, showing the amplitude distribution of spontaneous potentials (upper part), and of e.p.p. responses (lower part). The continuous curve in the lower part has been calculated on the hypothesis that the responses are built up statistically of units whose mean size and amplitude distribution are identical with those of the spontaneous potentials shown in the upper part. The expected number of failures is shown by arrows. Abscissae: scale units = mean amplitude of spontaneous potentials (0·875 mV).

(b) Summary of results of ten experiments similar to those illustrated in Figs. 11 and 12*a*. In each experiment, the mean amplitude of a large number of e.p.p. responses and spontaneous potentials were measured, and the proportion of "failures" and "successes" of e.p.p. responses were determined. The figure shows the validity of the relation given by eq. 1. Ordinate:

$$\log_e \frac{\text{number of impulses.}}{\text{number of e.p.p. failures}} \cdot \text{Abscissa:} \frac{\text{mean e.p.p. response}}{\text{mean amplitude of spontaneous potentials}}$$

The line corresponds to equality of these two estimates of m. (For further details see original paper.)

of successive e.p.p.'s) is accompanied by a recruitment of pre-synaptic units, and the converse happens during the slow progressive decline of a long-repeated response, known as "neuro-muscular depression."



(*a*) (*b*)

Fig. 13. This figure illustrates the statistical nature of neuro-muscular "facilitation." (From CASTILLO and KATZ.[87]) The nerve-muscle preparations were treated with solutions of raised Mg and reduced Ca content. The nerve was stimulated at a rate of 100 per second. The left part shows the beginning of the "recruitment" of units during a $\frac{1}{2}$ sec. period of stimulation; the right part, taken with slower speed, shows the progressive building up of the response. The upper records were obtained with external electrodes from a whole muscle and represent the statistical average response from a few hundred end-plates "in parallel." The lower portions were obtained with intracellular recording from single end-plates and show the discontinuous, quantal, nature of the growth of response. Nerve stimuli are indicated by dots. Note occasional spontaneous potentials on the superimposed "base-lines."

The changes which occur during "fatigue" of junctional transmission are of special interest. In the isolated nerve-muscle preparation of the frog, "fatigue" is associated with a large drop in ACh-output per nerve

impulse. But this fall is due to a reduction in the number of units, with little change in their individual sizes. The process is accompanied by a striking acceleration in the random firing of miniature e.p.p.'s.[87] It would appear that, during prolonged stimulation, some disorganization of the normal ACh-*release* mechanism occurs, forestalling any exhaustion, in the true sense, of the intracellular ACh supply. The situation may be different in the mammalian nerve-muscle system where the maintenance of junctional transmission depends on a steady supply of glucose, which is believed to ensure an adequate rate of intracellular synthesis and replenishment of ACh.[43], [89] One would expect this process to be graded at the molecular, rather than quantal, level, and it will therefore be interesting to see whether in a glucose-deprived mammalian preparation the unit of ACh-release is diminished in size. So far, all the factors which were found to influence the liberation of ACh appeared to operate statistically, on the size of the responding population rather than that of the individual units.

Some indirect evidence is available on the relation between the size of the presynaptic spike and the quantity of ACh liberated by it. It depends on the following observation. When the membrane potential of a nerve axon is raised "anelectrotonically," that is by flowing a steady inward current through the membrane, the amplitude and duration of an action potential which passes through the hyperpolarized region is increased.[35] When a moderate anelectrotonus is applied to the terminal portion of a motor axon, thus hyperpolarizing and increasing the size of the spike potential in the presynaptic endings, the e.p.p. —i.e. the effect of the transmitter recorded postsynaptically—becomes greater.[90] This must be an effect on the liberation of the transmitter, for the anelectrotonus is confined to the nerve structures and does not spread across the synapse. Furthermore, the miniature e.p.p.'s are not affected, in size or frequency, by an anelectrotnus of this relatively low intensity:[60] it appears therefore that the action potential in the nerve endings, being increased in amplitude and duration by the local hyperpolarization of the nerve membrane, is able to release a larger number of units of ACh.

While this observation indicates that the liberation of the transmitter varies with the size of the nerve impulse, there are other results which seem opposed to this view. Thus one might expect a fall in the ACh-output when the amplitude of the axon spike has been depressed by lowering the Na concentration, but HUTTER and KOSTIAL[74] failed to observe such a relation in the perfused ganglion. It is known, however, that in a medium of low Na content, the spike though reduced in amplitude increases in duration,[91] and it may be that this compensates the effect on the release of the transmitter substance.

That the efficacy of synaptic transmission depends on the amplitude

of the presynaptic axon spike, is suggested also by the findings of LLOYD,[92] namely that there is close correlation in time course between the "facilitation" of a spinal reflex, resulting from a period of preceding activity, and the spike height in the terminal parts of the afferent (presynaptic) nerve. The amplitude of the afferent impulse is increased in this case because of the prolonged local hyperpolarization ("positive after-potential") which follows a period of activity in these axons.

Apart from these observations, very little is known about the mechanisms responsible for "facilitation" and "depression" or "fatigue" at synapses. Recent studies at the myoneural junction indicate that these two phenomena originate at different stages in the chain of presynaptic events: thus "facilitation" occurs at a very early phase; it implies increased probability of ACh-release by a second nerve impulse, even when no ACh has been discharged by the first impulse.[87], [88] On the other hand, the delayed progressive decline of the transmitter output requires, as a necessary condition, that preceding nerve impulses release large quantities of ACh: "depression" is inconspicuous or absent, if the transmitter output has been reduced by Mg, while facilitation is very pronounced under these circumstances[87] (see also [93–98]).

To summarize the present information on the release of the neuromuscular transmitter, it can be said that liberation of ACh, spontaneously as well as after the arrival of a motor nerve impulse, takes place in discrete multi-molecular quanta whose probability of occurrence can be varied widely by experimental means but whose size is relatively constant. The mechanism by which a nerve impulse (or a steady depolarization[60]) releases such quanta of ACh is unknown, but it has been shown to depend upon the presence of external calcium ions, and to be inhibited by magnesium ions.

B. The mechanism of the ACh/receptor action

1. *Transfer of ACh from pre- to post-synaptic membrane*

Once the transmitter substance has been released from the nerve endings, the next stage is probably a simple process of diffusion. There is at any rate no reason, on our present evidence, to invoke anything more elaborate. The distances involved at the myoneural synapse are less than a micron, and the time required for the ACh-concentration to reach a peak at such short range is a small fraction of a millisecond.[64]

Recent investigations with the electron microscope[62] have revealed that the post-synaptic end-plate membrane is invaginated into a large number of deep folds (which gave the appearance of a regular pallisade-like structure in optical microscopy[99]). The farthest tips of the tubular folds are about $1{\cdot}\mu$ away from the pre-synaptic membrane. The contour of the nerve terminal does not follow the post-synaptic invaginations,

but is in intimate contact with their common base. The synaptic membranes of both nerve and muscle cells show three distinct layers, the outer ones being in direct apposition without detectable interspace.

One may wonder how ACh can escape through the synaptic layers into the perfusion fluid, or how an applied dose of ACh can reach the end-plate receptors presumably located in the deep folds of the post-junctional membrane. One might even feel tempted to argue that substances externally applied to, or collected from, the system have little do do with the "true" synaptic structures which look as though they are effectively sealed. Yet there can be no doubt that post-synaptic inhibitors of the curare type find the "true" neuroreceptors readily accessible and, having established blockage, can just as easily be washed away again. The answer to the apparent dilemma may be that the outer layers of the two cell membranes are made of a porous material which does not resist the passage of water molecules and aqueous solutes.

VOL. 6 1956

2. *The electromotive effect of acetylcholine*

The action of the transmitter on the postsynaptic membrane (henceforth called the "end-plate") has been investigated by two different methods: (i) by studying the changes in the electrical potential and conductance of the end-plate surface when a dose of ACh is applied to it from outside and (ii) by examining the "end-plate potential" (e.p.p.), i.e. the local depolarization which arises in the end-plate soon after the arrival of a nerve impulse and which is the first physical sign of the transmission process detectable with present methods.

If a muscle is placed in a solution which contains ACh, the end-plate region of the fibre becomes depolarized.[100], [101] Provided the concentration is not too high, the depolarization is maintained for many minutes. The sensitivity can be increased by previously incubating the tissue with a cholinesterase inhibitor. In a frog's sartorius muscle which has been pre-treated with prostigmine, bath concentrations of ACh of the order of 10^{-6} M lower the membrane potential at the end-plates sufficiently to initiate propagated spikes.[11] Much higher sensitivities are observed in chronically denervated muscles.[67]

Alternative methods of application are: intra-arterial injection,[48], [97], [102] small droplets,[51], [66], [67] or local micro-application by diffusion or electrophoresis.[8], [63], [64], [76], [103] In these cases, the concentration of ACh attained at the receptor site is not known, though theoretical estimates can be made. The method of rapid local application has the advantage of producing brief effects which can be repeated at frequent intervals, and in some respects these methods simulate more closely the impulsive action of the transmitter. It has already been mentioned that under suitable conditions a dose of less

than 10^{-15} M suffices to initiate muscle spikes, arising from an ACh-induced local depolarization.[8], [64] All these drug effects are strictly localized at the end-plate and cannot be obtained at remote points of the muscle fibre.

The mechanism of this local electrical change has been the subject of several recent investigations. The problem has various aspects: there is the question of the nature of the "receptor molecules" with which ACh is presumed to react. Secondly there is the unknown physico-chemical change which the membrane undergoes when ACh ions become attached to the receptor molecules. Finally, there are the resulting alterations in ionic permeabilities and flux rates which presumably determine the direction and magnitude of the electric potential change. The remaining sections of this review are devoted mainly to this last aspect of the "postsynaptic problem" on which a considerable amount of information has recently been collected. With regard to the properties of the receptor molecules and the chemical or structural changes in the end-plate membrane, other articles should be consulted.[11], [12], [19] It should be emphasized that the discussion is concerned with the specific case of neuro-muscular transmission and no generalization is possible from one type of synapse to another: for example, the transmission of the vagal nerve influence in the heart is also mediated by ACh, but its reaction with the receptors in this case leads to an increase of the membrane potential, not a depolarization.[104]

Does ACh act by inducing a permeability change or by specific penetration of the membrane? The first question that can be posed is whether ACh depolarizes the end-plate by making the membrane more permeable to other ions, or whether ACh^+ ions are capable of penetrating the membrane, by some specific process, at a very high rate and depolarize the cell by transferring their own charge into its interior.[105] The latter process would be analogous to a mechanism suggested by BARNES and BEUTNER[106] as a result of their study of electrical phenomena at oil/water interfaces. The possibility of a specific penetration of the end-plate by ACh^+ ions was considered by FATT[101] in an attempt to explain the residual depolarizing action observed in a Na-free solution. More recent experiments, however, by FATT and KATZ[7] and by NASTUK[8], [83] have ruled out this type of mechanism or, at least, relegated it to an unimportant place.

The decisive finding was that the transfer of electrical charge across the end-plate, which occurs during the local depolarization, is much greater than the total charge carried by the quantity of ACh released by an impulse[7] and, under suitable conditions, even by electrophoretically applied ACh^+.[8] It appears, therefore, that the action of ACh^+ brings about an amplified flow of current through the end-plate, and it does so presumably by increasing the permeability of the

membrane to other ions which are present in large quantities in the surrounding media.

The question arising is what kind of permeability change takes place. It might be a specific increase of the sodium conductance, such as occurs during the ascending phase of the axon spike.[83] Or it may be a less specific change, involving various ions and not discriminating sharply between sodium and potassium. Taking a more extreme view, FATT and KATZ[7], [107] suggested that ACh may "short-circuit" the membrane, i.e. create aqueous channels through which small ions can pass without specific distinction. The principal arguments for and

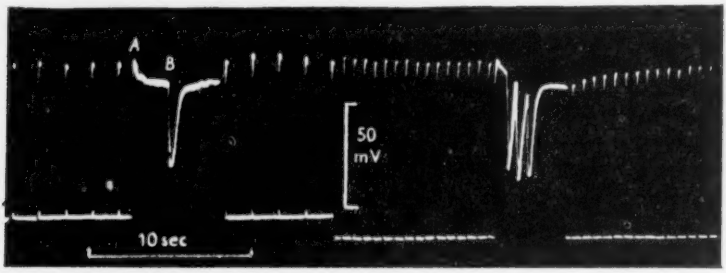


Fig. 14. The effect of acetylcholine on the electric resistance of the end-plate membrane, treated with isotonic K_2SO_4. (From CASTILLO and KATZ.[77]) The cell membrane is completely depolarized in this medium, but by passing current through a second intracellular electrode (cf. Fig. 4), the membrane potential can be electrotonically displaced from zero. The current was applied at first in the form of brief pulses, and at *A* was maintained to produce a prolonged change of the membrane potential. At *B*, a quantity of ACh was momentarily applied from a close external pipette (cf. Fig. 4*c*). The effect is a prompt decline of the electrotonic potential, showing that ACh reduces the resistance of the end-plate membrane even in an isotonic K_2SO_4 bath. The procedure was later repeated, this time applying three successive charges of ACh.

against these different views were derived from a study of the end-plate potential and will be discussed in the following section. Some subsidiary evidence, however, may be mentioned here.

On all these hypotheses the observed potential change would be due largely to a net entry of sodium, because of the relatively high external concentration and mobility of these ions and the large inward driving force normally acting on them. A simultaneous increase of potassium and chloride conductance would influence the final level, but not the initial rate of development of the ACh-potential. When sodium is withdrawn from the bath solution and replaced by sucrose, the depolarizing effect of ACh is reduced or even abolished.[76], [77], [83] The experiment is not entirely satisfactory because it is possible that the unavoidable lowering of the ionic strength of the solution desensitizes the end-plate receptors. In similar experiments on the axon or muscle fibre membrane, such difficulties have been overcome by

substituting apparently "inert" quaternary ammonium ions for sodium.[35] At the end-plate, none of these compounds can be used; they either act like ACh or inhibit like curare. The small and rather irregular depolarization which ACh may still produce in a Na-free medium is of doubtful significance; it possibly indicates that efflux of intra-cellular anions can contribute to the ACh-effect.

Concerning the role of potassium and chloride, simple potential measurements are not likely to tell us whether these ions are involved. For if they are in electro-chemical equilibrium between cell and surroundings,[108] then a change of permeability to K or Cl is not going to alter the membrane potential, until it has been deflected by other means from the equilibrium level. Thus, the absence of a depolarization in a Na-free solution does not imply that ACh fails to raise the permeability to, e.g. K ions. This particular case has been examined by immersing the muscle in an isotonic solution of K_2SO_4.[77], [78] In this medium, the cell membrane is completely depolarized, and application of ACh does not produce a potential change in either direction. But if the membrane potential is displaced from zero by the passage of a transverse current, renewed application of ACh reduces the p.d. built up across the end-plate membrane. This experiment shows that in spite of the abnormal conditions (depolarization, electric inexcitability, absence of Na) ACh still acts on the end-plate receptors and increases the postsynaptic membrane conductance. The change is presumably accompanied by an increased flux of K ions which are present in high concentration on either side of the cell membrane. It has already been said that in the same type of experiment continuance was detected of the quantal discharge of ACh from the nerve endings producing intermittent spontaneous reductions of the post-junctional membrane resistance.

It appears from these experiments that ACh raises both the Na and K conductance of the end-plate and possibly also its permeability to other free ions. ACh normally depolarizes by allowing Na ions to enter, but unlike the process which leads to the peak of the axon spike, the conductance change at the end-plate is not Na-selective.

3. *The nature of the end-plate potential*

The experiments hitherto described show that ACh, when applied externally, depolarizes the end-plate. An analogous phenomenon is observed during neuro-muscular transmission, namely a transient intense local depolarization of the end-plate region, which follows the arrival of a nerve impulse and precedes the impulse in the muscle fibre. This effect has been called the "end-plate potential" (e.p.p.). Apart from its discrete timing it possesses a number of characteristics which distinguish it from the muscle spike. Normally, the e.p.p. rises above

VOL
6
195

the excitation threshold of the muscle membrane and so leads quickly to the initiation of a propagated action potential.[71] This complicates the analysis of the e.p.p., because its further time course is obscured by the superposition of the muscle spike. To simplify matters, the spike may be removed experimentally by reducing the e.p.p. to subthreshold size. A simple way of obtaining this result is to use curare which specifically interferes with and reduces the postsynaptic depolarizing action of the nerve-muscle transmitter (and also that of applied ACh[47], [67], [70], [100]).

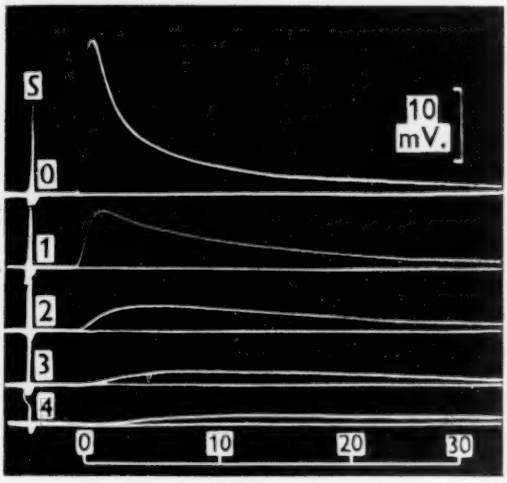


Fig. 15. The end-plate potential in a curarized muscle fibre. (From FATT and KATZ.[118]) The moment of stimulation of the motor nerve is marked by the artifact *S*. The response is an e.p.p. which disperses electrotonically along the fibre surface. The numbers on the records show the distance, in millimetres, between the end-plate and the point of insertion of the recording electrode. Time scale: milliseconds.

When the e.p.p. is examined in such a curarized muscle fibre, it is found to resemble in some aspects a subthreshold electric signal: it spreads a small distance along the muscle membrane, and because of the defective cable-properties of the fibre (see p. 126) is rapidly "blurred" and attenuated.[7], [70] These changes can be accurately predicted from the separately determined "cable constants" of the muscle fibre, i.e. its distributed capacity and resistances. The result of such an analysis[7] has shown that after the arrival of a motor nerve impulse, the end-plate undergoes a very short-lived disturbance (a transient increase of its ion permeability lasting a few milliseconds) after which the membrane properties return to their resting condition, and the potential change gradually dissipates itself, spreading

electrotonically and leaking across the fibre membrane in the manner of an ordinary subthreshold electric pulse.

It is pertinent that the duration of the "active phase," i.e. of the period during which the transmitter acts upon the end-plate and disturbs its ionic permeability, can be greatly lengthened by lowering the temperature[70] or treating the tissue with a cholinesterase inhibitor[7], [51] (a striking example being shown in Fig. 5). Evidently, one of the factors which limit the period of transmitter activity at the myoneural junction, is the rate of enzymic hydrolysis of ACh.

From measurements of the e.p.p. and of the electric capacity of the cell membrane, it has been possible to determine the quantity of electric charge which flows through the end-plate during the period of transmitter activity.[7] In a curarized fibre, it amounts to about 8×10^{-10} coulomb; during normal transmission, this quantity is increased by a factor of 3–4, and after treatment with a cholinesterase inhibitor by a further factor of up to 50 times (about 10^{-7} coulomb). The transfer of charge must be associated with ionic current across the end-plate membrane, and a quantity of 10^{-7} coulomb requires a net flux of at least I $\mu\mu$M of univalent ions. This is probably more than the total ionic content of the nerve terminals; it means, in effect, that the local current which builds up the e.p.p. cannot have been derived directly from the nerve endings (e.g. from their action currents, or the flow of transmitter ions), but must be due to a resistance breakdown in the post-synaptic membrane, thus allowing a vastly amplified pulse of ionic current to discharge through the end-plate surface.

The e.p.p. is, in its essential properties, analogous to the "ACh-potential," that is to the local depolarization produced by an applied dose of ACh. What differs between the two situations is mainly the distribution in time and space of the ACh-concentration. It has not been possible, with a micropipette, to approach the end-plate receptors within a comparable range and to achieve close contact similar to that of the synapse. The artificially produced ACh-potentials are therefore unavoidably of slower time course than the normal e.p.p.[8], [64] It is true that the electrophoretic method allows one to release ACh in an interval as brief as the nerve impulse, but the diffusion path between pipette and end-plate delays and attenuates the effect and makes it last much longer than the impulsive short-range action of the neuromuscular transmitter. The time course of the e.p.p., apart from its brief rising phase which depends on the short life of the transmitter, is determined by the resistance and capacity of the resting fibre membrane; it represents a passive return of the electric charge which has been initially displaced by the transmitter action. The "ACh-potential," however, represents in its time course the gradual waxing and waning of the ACh-concentration at the receptor site; this is generally such a

slow process that the resulting potential change is not greatly affected by the electrical time constant of the fibre membrane.

If we compare, finally, the e.p.p. with the action potential in a nerve or muscle fibre, we find the following important differences (Table 1);

TABLE 1

Properties of Spike and End-plate Potential

	Spike	*E.P.P.*
Initiated by . .	(Outward) electric current	Acetylcholine
Changes of membrane conductance: (*a*) During rising phase	Specific increase of Na-permeability	Increased permeability to Na *and* K (and other ions?)
(*b*) During falling phase	Specific increase of K-permeability	No increase of ion permeability above resting condition (i.e. "passive" decay)
Equilibrium potential of active membrane	Na-potential (approx. 50 mV, inside positive)	"Diffusion potential" (approx. 15 mV, inside negative)
Other distinguishing features	Regenerative ascent; followed by refractory period	No evidence for regenerative action or refractoriness

The terms "active" and "passive" are used here to distinguish processes which involve a reactive change in the cell properties in response to a stimulus (e.g. an increased membrane permeability) from those in which no such change appears to be involved (e.g. simple electrotonic potentials). Some authors confine the term "active" to reactions which are positive (regenerative, like the increase of Na-permeability during depolarization); but in the present context negative reactions (like the adaptive increase of K-permeability) are included.

In other recent discussions, the term "active" cell process has been defined as one in which "work" is done by the cell in moving substances against a gradient of physical or chemical forces. On this terminology, the brief ionic exchange during the nerve impulse would be classified as a passive process involving merely the release of stored energy.

the action potential rises by a regenerative electrical process coupled with a specific increase of Na-permeability (p. 125); its falling phase is associated with, and greatly speeded up by, a specific rise of K-permeability. By contrast, the ascending phase of the e.p.p. is due to a more indiscriminate rise of ion permeability (Na and K being involved simultaneously) brought about by the local action of ACh, while the fall of the e.p.p. occurs after the ionic conductances have been restored to their normal level, and is controlled by the time constant of the resting fibre membrane.

Although it is true that, in both cases, entry of external Na ions is normally the principal means by which the membrane becomes depolarized, the situation at the end-plate differs from that during the fibre spike in two respects: (*a*) the electrical change shows no signs of being regenerative, (*b*) increased efflux of K ions occurs simultaneously and opposes or prevents a reversal of the membrane potential.[7], [109], [110]

On the basis of these findings, it is now much easier to interpret the complicated phenomena which are observed at the normal nerve-muscle junction where the e.p.p. is large and gives rise to, and interacts with, the action potential of the muscle fibre.

An impulse is initiated in the muscle fibre when the local depolarization of the cell membrane exceeds some 30 to 40 mV. It makes no difference whether this depolarization was brought about by the neuromuscular transmitter, or by applied ACh, at the end-plate, or by suitable electric current passing through any point of the fibre.[7], [107]

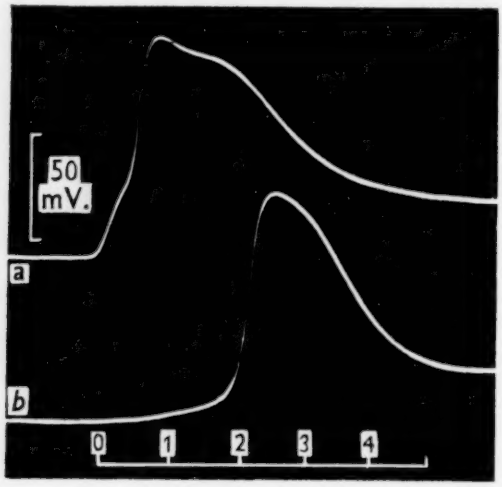


Fig. 16. Action potentials at, and off, the end-plate in a single muscle fibre. (From FATT and KATZ.[118]) The nerve was stimulated at the beginning of the time base, and the membrane action potential was recorded: (*a*) at the end-plate and (*b*) 2·5 mm away. Time scale: milliseconds. Note the delay of the conducted spike, and the complex shape of the action potential at the end-plate (for a detailed description see FATT and KATZ[7] and NASTUK[83])

The action potential which arises when this threshold level is exceeded has, usually, an amplitude of about 120–130 mV (the membrane potential at the peak being about 35 mV, negative outside) and propagates at a constant velocity to the tendon ends of the fibre.

However, the amplitude of the spike is *less* at the end-plate, during normal nerve-muscle transmission, than at other points of the fibre. The active-membrane potential is only about 22 mV at the end-plate, instead of 35 mV elsewhere.[7], [83] If the impulse is set up by *direct* electric stimulation of the muscle fibre, this reduction at the end-plate is not observed, and the amplitude of the spike is the same at all points of the fibre. There are other characteristic differences between the "neuro-" and "myo-genic" action potentials of the end-plate region;

thus the former has a complex shape, with the potential wave rapidly falling from its diminished crest to a lower plateau from which the final decline to the base-line proceeds after a distinct delay.

The properties of this complex action potential can be explained by the simultaneous occurrence in adjacent parts of the cell membrane of two different types of electric activity, e.p.p. and spike. The spike is due to a very brief and specific increase of Na-permeability, tending to

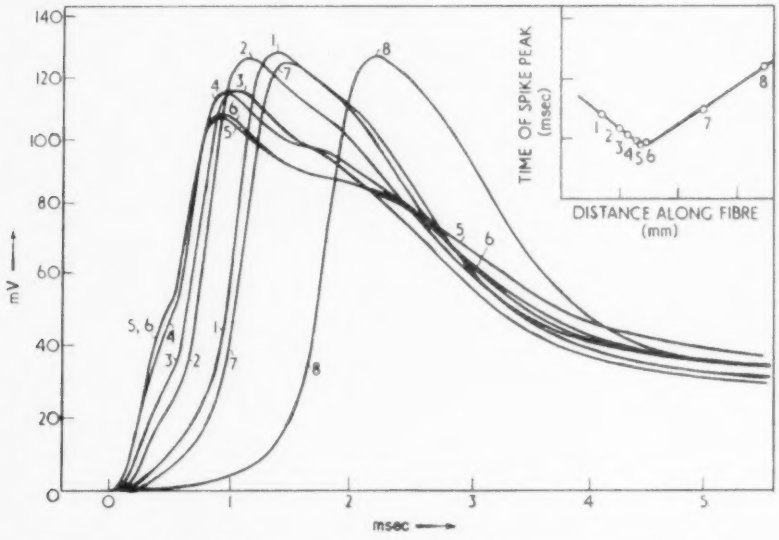


Fig. 17. The initiation of a muscle action potential. (From FATT and KATZ.[7]) This figure illustrates the transition of the electric response from the end-plate region to the rest of the muscle fibre. The microelectrode was moved along the fibre and inserted at the distances shown in the inset. Note the gradual changes in the shape and delay of the action potential. Inset: the time of the peak of the spike is plotted against distance showing a propagation velocity of about 1·4 m/sec. in both directions from the nerve-muscle junction (at positions 5 and 6).

displace the membrane potential towards the equilibrium level for the existing concentration difference of sodium ions (i.e. about 50 mV, negative outside). The e.p.p. is due to a slightly more prolonged unspecific, or less specific, increase of ion permeability, involving Na, K and perhaps also other ions, and tending to *depolarize* the membrane, or to displace its potential towards the low level of a diffusion potential between cell contents and surroundings (about 10–20 mV, negative inside). The two changes occur at adjacent but different spots of the membrane, and their electric interaction is, roughly speaking, similar to that between two batteries of different voltages and different internal resistances which have been placed in parallel. The amplitude of the

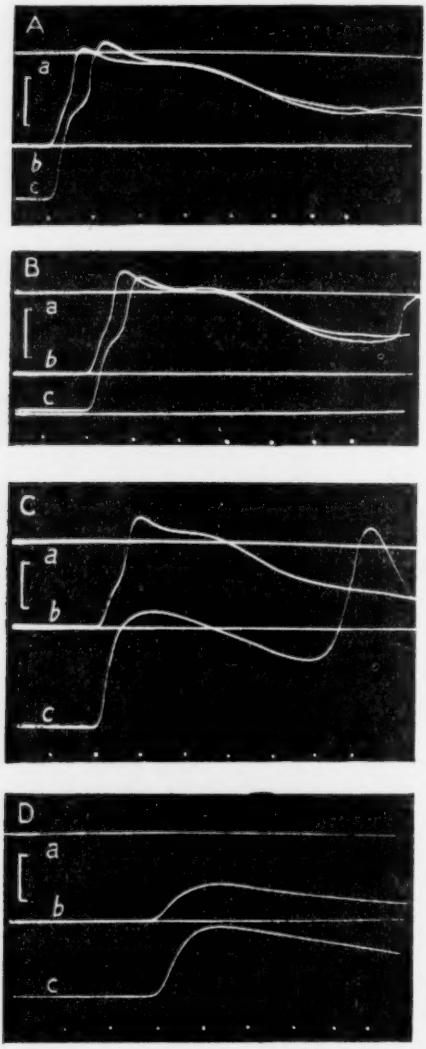


Fig. 18. The effect of an increase in the post-synaptic membrane potential (local analectrotonus) on the size of the e.p.p. (From FATT and KATZ.[7]) (*a*) zero p.d. across membrane; (*b*) normal resting potential; (*c*) resting potential has been increased by applied inward current. *A–D*: Four different fibres. In *C*, transmission is blocked by the inward current, while in *D* transmission had failed at the normal level of the resting potential. Voltage scale: 50 mV, time marks: milliseconds.

"sodium-spike" in the surrounding fibre membrane is reduced by the simultaneous presence of a "short-circuit" at the end-plate which tends to keep the membrane depolarized and to prevent its potential from attaining a high reversed level.

In addition to the findings discussed above, further evidence has accumulated in support of this interpretation. Thus, the e.p.p. increases in proportion to the resting potential if the latter is raised locally by an anodic current (Fig. 18[7]). This is to be expected from the "short-circuit" theory according to which the transmitter produces a constant

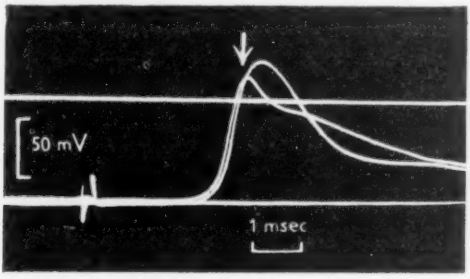


Fig. 19. Intra-cellular recording from a single end-plate. This figure illustrates an experiment in which a muscle was stimulated directly (by electric current) and, in addition, a nerve impulse was made to arrive at the junction during the passage of the directly excited muscle spike. The larger of the two superimposed action potentials is a simple, directly excited, spike. The arrow indicates the commencement of the "neural" action. The release of the transmitter, at his moment, causes the action potential to *decline* towards a transient plateau some 10 mV below the zero-potential line. (From CASTILLO and KATZ.[110])

degree of depolarization (i.e. reduces the membrane potential by a constant fraction), if one ignores the presence of a limiting liquid junction potential.

More recently, it has been possible to release the transmitter at various phases of a "myogenic" spike.[109], [110] In these experiments, it was found that the neuro-muscular transmitter tends to re-set the membrane potential and to shift it towards an equilibrium level of about 10–20 mV, negative inside. This occurs irrespectively of the level of the action potential at which the ACh-effect intervenes. If the transmitter is released at a lower level of the spike it causes ionic current to flow inward through the end-plate and produces an e.p.p. which adds to that part of the spike. If the transmitter is released at a higher level (near the peak of the action potential), ionic current flows outward and produces a deflection of opposite direction, *subtracting* from the muscle spike. The reversal point is close to a level estimated for a free-diffusion, or liquid junction, p.d. between fibre contents and surroundings, and thus supports the short-circuit theory.[110]

Finally, it has been shown that the release of the transmitter is followed by a large local increase of the membrane conductance, greater than that occurring during the peak of a "myogenic" spike.[110] This again agrees with the short-circuit hypothesis according to which the transmitter produces an additional leakage path in parallel with the

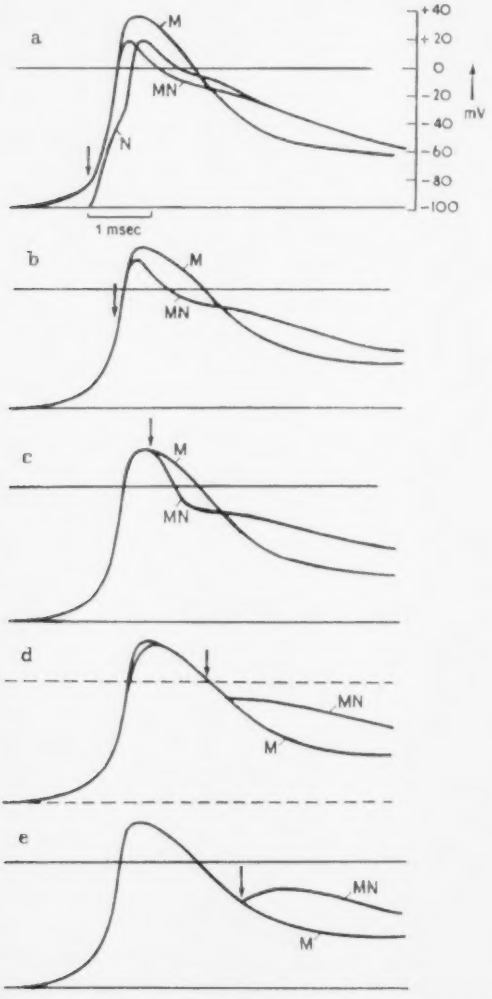
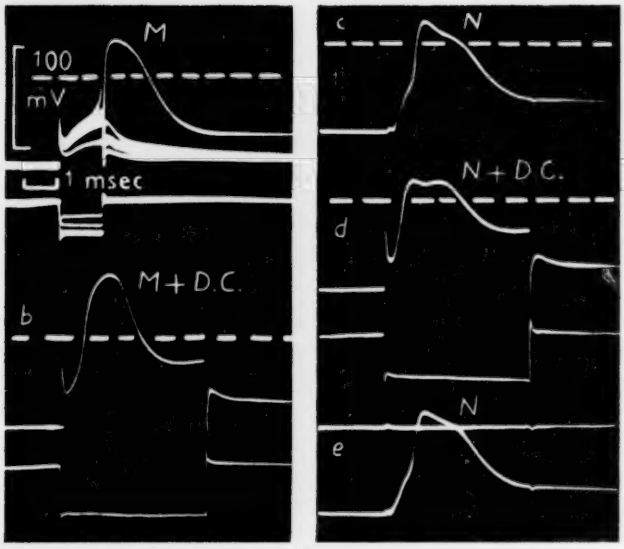
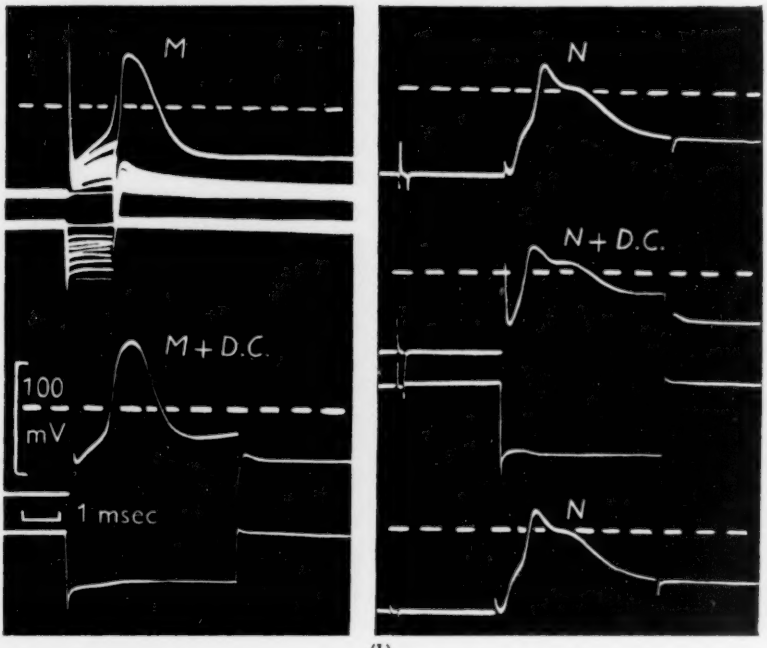


Fig. 20. Superimposed tracings of records from a single end-plate, showing the interaction, at various phases, between nerve-muscle transmitter and directly excited muscle spike. (From CASTILLO and KATZ.[110]) *M*: response to direct stimulation. *N*: response to a nerve impulse. *MN*: modified response due to combined *M* and *N* stimuli. The arrows indicate the start of the *N*-responses. Note: the effect of the nerve-muscle transmitter is to displace the membrane potential towards a level about 10–20 mV below zero.



(*a*)



(*b*)

Fig. 21. Membrane resistance at the active motor end-plate. (From CASTILLO and KATZ.[110]) The resistance was measured by passing a current pulse through the end-plate membrane and observing the added p.d. *A* and *B*: examples from two experiments. *M* are responses to direct stimuli, *N* responses to nerve stimuli. *M* + *D.C.* and *N* + *D.C.*: in these records, an outward current (registered immediately below the action potential) was passed through the membrane during the spike. This produced an addition to the spike potential which was greater during *M* + *D.C.* than during the corresponding *N* + *D.C.* record. It appears therefore that the neuro-muscular transmitter increases the membrane conductance of the end-plate above the level reached during a directly initiated muscle spike.

normal channels of ionic flux within the fibre membrane. Moreover, the increase of the end-plate conductance is of the right order of magnitude to account (*a*) for the observed rate of depolarization during the rise of the e.p.p. and (*b*) for the diminution of the spike amplitude occurring in the surrounding fibre membrane. It should be emphasized that the change in "specific conductance" (i.e. conductance per unit surface of membrane) must be much larger at the end-plate, during neuro-muscular transmission, than in the surrounding fibre during the propagation of the impulse, for, while the absolute magnitude of the two changes is of the same order, the effective surface area in which the change occurs is highly localized in the case of the e.p.p., but distributed along the fibre in the case of the impulse. An exact comparison, however, is not possible at present, and it will be necessary to take into account the local surface expansion in the postsynaptic membrane which has been described by ROBERTSON.[62]

The mechanism of the transmitter action seems well suited to the particular requirements of the neuro-muscular junction, where the impulse arrives in relatively minute axon branches and excites, without fail or delay, a much larger effector cell. A high amplification of ionic current must be achieved at the point where the impulse is to be transferred from the nerve endings to the muscle fibre. This is brought about by the action of ACh which seems tantamount to a transient "puncture" of the receptor membrane. This effect rapidly discharges the surrounding fibre membrane, starting off a renewed wave of electric excitation which travels quickly along the muscle fibre and sets in motion the contractile mechanism at every point.

C. The removal of the transmitter from its site of action

Three different factors have been discussed which may contribute to the removal of ACh from the end-plate receptors: (*a*) hydrolysis by a localized enzyme system, (*b*) diffusion into the interstitial tissue spaces and, possibly (*c*) resynthesis or, more generally, restitution to the presynaptic terminals.

That a specific cholinesterase is responsible, at least in part, for the rapid inactivation of the transmitter is indicated by several facts: e.g. by the high local concentration of the enzyme[49] which can be demonstrated histochemically at the motor end-plate,[50] and by the striking prolongation of the transmitter effect under the influence of specific esterase inhibitors.[7], [48], [51], [52] The relatively slow subsidence of the transmitter action in these experiments may be due to other causes, principally diffusion into the interspaces.

It is of some interest that the duration of the rising phase of the e.p.p. (indicating the period of persistent activity of the transmitter) has a high positive temperature coefficient:[70] this might conceivably be

ascribed to the influence of temperature on the rate of enzymic hydrolysis, and it will be worth examining whether the large temperature coefficient becomes significantly reduced when the enzyme activity has been stopped by a chemical inhibitor.

That diffusion plays a part in removing ACh from the end-plate seems obvious and is of course substantiated by the fact that ACh can be collected in the perfusion fluid provided it is protected from hydrolysis by an inhibitor like eserine.[48] In theory, inactivation by diffusion from a highly localized source could be a very rapid process; but too little is known about the special conditions of the synapse to permit any accurate predictions.

It is still uncertain whether a direct return of the transmitter to its presynaptic source takes place. There is indirect evidence that such a process of local restitution may occur, and it might explain certain discrepancies between observed results and the *prima facie* expectations of the transmitter theory. Thus, the amount of ACh which is released into the *eserinized* fluid from a stimulated sympathetic ganglion is greater than the quantity of *choline* which can be collected in the absence of an esterase inhibitor (no significant amounts of ACh are then obtainable[111]). PERRY[111] has suggested that there is a local recovery process by which choline (but not acetylcholine) is taken up again and re-synthetized in the nerve ending. On this view, the function of the esterase would be not only to inactivate the transmitter substance, after it has produced an impulsive effect, but to provide the first essential step of a recovery reaction.

There is another way in which ACh can become "inactivated." It has been observed by several authors[101], [103] and particularly clearly by THESLEFF[112] that the local potential change produced by a large dose of ACh gradually declines, even though the ACh concentration in the vicinity of the end-plate is maintained at a high level. It appears that the end-plate receptors can become "refractory" to the depolarizing substance. During the state of "refractoriness," further increases in ACh-concentration, either by external application or by added nerve impulses[112] may become totally ineffective. Removal of the drug is followed by very slow, and often incomplete, recovery of the end-plate sensitivity.

Although this effect is artificial and not likely to occur in the course of normal nerve activity, it is of great potential interest for an understanding of the drug/receptor reaction. The phenomenon might be due to two different mechanisms: (*a*) there may be a gradual change in the end-plate membrane analogous to the delayed increase of K- and decrease of Na-permeability in a depolarized axon membrane. In this case, the resting potential would slowly become re-established, but the membrane resistance would remain low. Such a process should be

revealed, even after subsidence of the ACh-potential, by a shortening of the membrane time constant and a local depression, or blockage, of directly excited ("myogenic") impulses. Neither of these predictions agrees with experimental results (unpublished observations by W. BURKE and B. KATZ), and this points to the alternative explanation. (*b*) The character of the receptor response to ACh may change from the normal to a "curare-type" of reaction which is, of course, merely another way of saying that the ACh-receptor complex becomes eventually "inert" and fails to maintain or produce a high ionic permeability of the end-plate membrane.

Neuro-muscular, post-synaptic, blocking agents have been divided into two distinct groups:[102], [113] those which depolarize, and in large doses rendor the muscle fibre locally inexcitable by excessive maintained depolarization; and those, like curare, which produce "competitive inhibition," i.e. combine with the receptor to form an inert complex which does not change the electrical properties of the membrane. It now seems that in many instances the "depolarization block" is only a transient event and ultimately passes over into "curare-like" inhibition.[112], [114], [115] The fact that the response to a given substance can change from one type to the other cuts across the hitherto accepted classification and indicates that the receptor properties themselves are liable to vary as a result of previous activation.[115]

V. CONCLUSIONS

Recent work on the nerve-muscle junction has confirmed the chemical transmission theory and provided further evidence on the intermediate steps of the process. In addition to the original findings of DALE and his colleagues, that acetylcholine is released from motor nerve endings and, if applied directly, can set up muscle impulses, there is now complementary evidence to show (*a*) that local currents do not spread across the synapse, (*b*) that small quantities of ACh increase the ionic conductance of the end-plate membrane and give rise to an intense local current flow, (*c*) that the efficacies of artificially applied and naturally released amounts of ACh are not incongruous, and (*d*) that the specific chemoreceptors of the end-plate are attached to the *external* surface of the post-synaptic membrane, which faces the nerve terminals.

Comparatively little has been revealed about the mechanism by which nerve impulses liberate, or facilitate the release of, acetylcholine. ACh exists in "bound" form inside the nerve terminals, possibly attached to the electron-microscopic granules observed by ROBERTSON,[62] and it appears that the process of "release" requires a simultaneous lowering of more than a single diffusion barrier. The discharge of ACh, spontaneously or in response to an impulse, occurs in "quantal," i.e. multi-molecular, units. The *size* of this quantum seems to be

constant under a variety of experimental conditions; it may conceivably be determined by the "storage capacity" of individual intracellular ACh-carriers. The total *number* of ACh-quanta which is released at any one junction by a nerve impulse increases with the calcium concentration and decreases with rising magnesium concentration in the surrounding medium.

The process by which ACh is transferred, after its release, to the site of the post-synaptic receptors is not known. However, the distances are less than one micron, and simple diffusion at this close range would be both fast and effective. This view is supported not only in theory, but by experiments with electrophoretic microapplication of ACh.[8], [64], [83]

Very little is known about the chemical properties of the postulated receptors and the immediate physico-chemical changes resulting from the formation of an ACh/receptor complex. The effect of this reaction, however, on the electrical properties of the membrane has been studied and found to involve a large increase of permeability, to sodium and potassium, and possibly other free ions. This effect differs from the specific and strictly separate, successive changes in Na and K permeability which occur during electric excitation of the axon or muscle membrane.

There are other reasons for believing that the ACh-action is a different and independent event; thus, it has been shown that the transmitter, or external ACh, increases the ion permeability of the membrane even when it is released at the peak of the action potential, or during a state of persistent complete depolarization of the muscle fibre, when electric excitability has been eliminated.

Apparently, we must look upon the "chemo-receptors" of the end-plate as different molecular entities, separate from those responsible for the ionic exchanges during electric excitation. As both types of structures, or molecules, are constituents of the same cell membrane and situated side-by-side, simultaneous local conductance changes during e.p.p. and spike are additive (as found by CASTILLO and KATZ,[110]) and the membrane potential at the active end-plate is the resultant of two different sources of e.m.f. connected "in parallel."[107]

VOL. 6 1956

REFERENCES

(1) HODGKIN, A. L. and HUXLEY, A. F.; *Nature*, Lond. 144 (1939) 710
(2) HODGKIN, A. L. and HUXLEY, A. F.; *J. Physiol.* 104 (1945) 176
(3) CURTIS, H. J. and COLE, K. S.; *J. cell. comp. Physiol.* 19 (1942) 135
(4) GRAHAM, J. and GERARD, R. W.; *J. cell. comp. Physiol.* 28 (1946) 99
(5) LING, G. and GERARD, R. W.; *J. cell. comp. Physiol.* 34 (1949) 382
(6) NASTUK, W. L. and HODGKIN, A. L.; *J. cell. comp. Physiol.* 35 (1950) 39
(7) FATT, P. and KATZ, B.; *J. Physiol.* 115 (1951) 320
(8) NASTUK, W. L.; *Fed. Proc.* 12 (1953) 102

(9) BROCK, L. G., COOMBS, J. S. and ECCLES, J. C.; *J. Physiol.* 117 (1952) 43
(10) WOODBURY, J. W. and PATTON, H. D.; *Cold Spr. Harb. Symp. quant. Biol.* 17 (1952) 185
(10a) KUFFLER, S. W. and VAUGHAN WILLIAMS, E. M.; *J. Physiol.* 121 (1953) 289
(11) FATT, P.; *Physiol. Rev.* 34 (1954) 674
(12) RIKER, W. F.; *Pharmacol. Rev.* 5 (1953) 1
(13) BARLOW, R. B. and ING, H. R.; *Brit. J. Pharmacol.* 3 (1948) 298
(14) ING, H. R.; *Science* 109 (1949) 264
(15) WELSH, J. H.; *Amer. Scientist* 38 (1950) 239
(16) HUNT, C. C. and KUFFLER, S. W.; *J. Pharmacol.* 98 (1950) 96
(17) ROSENBLUETH, A.; *The Transmission of Nerve Impulses at Neuro-effector Junctions and Peripheral Synapses.* New York, 1950
(18) NACHMANSOHN, D.; *Ann. N.Y. Acad. Sci.* 47 (1946) 395
(19) NACHMANSOHN, D.; in *Le muscle—Etude de biologie et de pathologie* (1950) 121
(20) HODGKIN, A. L.; *Biol. Rev.* 26 (1951) 339
(21) HODGKIN, A. L. and HUXLEY, A. F.; *J. Physiol.* 117 (1952) 500
(22) HODGKIN, A. L. and KEYNES, R. D.; *Symp. Soc. Exp. Biol.* 8 (1954) 423
(23) HARRIS, E. J. and BURN, G. P.; *Trans. Faraday Soc.* 45 (1949) 508
(24) KEYNES, R. D.; *Arch. Sci. physiol.* 3 (1949) 165
(25) HODGKIN, A. L. and KEYNES, R. D.; *J. Physiol.* 119 (1953) 513
(26) HARRIS, E. J.; *J. Physiol.* 124 (1954) 248
(27) COLE, K. S. and HODGKIN, A. L.; *J. gen. Physiol.* 22 (1939) 671
(28) HODGKIN, A. L. and RUSHTON, W. A. H.; *Proc. Roy. Soc.* B. 133 (1946) 444
(29) KATZ, B.; *Proc. Roy. Soc.* B. 135 (1948) 506
(30) HODGKIN, A. L. and HUXLEY, A. F.; *J. Physiol.* 121 (1953) 403
(31) BOGUE, J. Y. and ROSENBERG, H.; *J. Physiol.* 82 (1934) 353
(32) RUSHTON, W. A. H.; *Proc. Roy. Soc.* B. 124 (1937) 210
(33) KATZ, B.; *Electric Excitation of Nerve.* Oxford, 1939
(34) SCHAEFER, H.; *Elektrophysiologie*, vol. 1. Vienna, 1940
(35) LORENTE DE NO, R.; *A Study of Nerve Physiology*, vols. 1 and 2. New York, 1947
(36) HODGKIN, A. L.; *J. Physiol.* 90 (1937) 183, 211
(37) LORENTE DE NO, R.; *J. Neurophysiol.* 2 (1939) 402
(38) KATZ, B. and SCHMITT, O. H.; *J. Physiol.* 97 (1940) 471
(39) ARVANITAKI, A.; *J. Neurophysiol.* 5 (1942) 89
(40) LANGLEY, J. N.; *J. Physiol.* 33 (1905) 374
(41) LANGLEY, J. N.; *J. Physiol.* 36 (1907) 347
(42) FELDBERG, W.; *J. Physiol.* 101 (1943) 432
(43) FELDBERG, W.; *J. Physiol.* 103 (1945) 367
(44) FELDBERG, W.; *Physiol. Rev.* 25 (1945) 596
(45) LOEWI, O., HAGEN, R., KOHN, H. and SINGER, G.; *Pflüg. Arch. ges. Physiol.* 239 (1938) 430
(46) LOEWI, O. and HELLAUER, H.; *Pflüg. Arch. ges. Physiol.* 240 (1938) 449
(47) DALE, H. H., FELDBERG, W. and VOGT, M.; *J. Physiol.* 86 (1936) 353
(48) BROWN, G. L., DALE, H. H. and FELDBERG, W.; *J. Physiol.* 87 (1936) 394
(49) MARNEY, A. and NACHMANSOHN, D.; *J. Physiol.* 89 (1937) 359
(50) KOELLE, G. B. and FRIEDENWALD, J. S.; *Proc. Soc. exper. Biol., N.Y.* 70 (1949) 617
(51) ECCLES, J. C., KATZ, B. and KUFFLER, S. W.; *J. Neurophysiol.* 5 (1942) 211
(52) ECCLES, J. C. and MACFARLANE, W. V.; *J. Neurophysiol.* 12 (1949) 59
(53) NACHMANSOHN, D.; *Abstr. XVIII Internat. Physiol. Congr.* (1950), p. 371
(54) GRUNDFEST, H., NACHMANSOHN, D., KAO, C. Y. and CHAMBERS, R.; *Nature* Lond. 169 (1952) 190

VOL
6
195

(55) BIEDERMANN, W.; *Elektrophysiologie.* Jena, 1895

(56) JASPER, H. H. and MONNIER, A. M.; *J. cell. comp. Physiol.* 11 (1938) 259

(57) KUFFLER, S. W.; *Arch. Sci. physiol.* 3 (1949) 585

(58) FATT, P. and KATZ, B.; *J. Physiol.* 117 (1952) 109

(59) CASTILLO, J. DEL and KATZ, B.; *J. Physiol.* 123 (1954) 7*P*

(60) CASTILLO, J. DEL and KATZ, B.; *J. Physiol.* 124 (1954) 586

(61) ACHESON, G. H.; *Fed. Proc.* 7 (1948) 447

(62) ROBERTSON, J. D.; *J. biophys. biochem. cytol.* (1955). The ultrastructure of a reptilian myoneural junction. In preparation.

(63) CASTILLO, J. DEL and KATZ, B.; *J. Physiol.* 125 (1954) 16*P*

(64) CASTILLO, J. DEL and KATZ, B.; *J. Physiol.* (1955). 128 (1955) 157

(65) CARSLAW, H. S. and JAEGER, J. C.; *Conduction of heat in solids.* Oxford, 1947

(66) BUCHTHAL, F. and LINDHARD, J.; *J. Physiol.* 90 (1937) 82*P*

(67) KUFFLER, S. W.; *J. Neurophysiol.* 6 (1943) 99

(68) FATT, P. and KATZ, B.; *Cold Spr. Harb. Symp. quant. Biol.* 17 (1952) 275

(68a) KUFFLER, S. W.; *Fed. Proc.* 7 (1948) 437

(69) HODGKIN, A. L. and KEYNES, R. D.; in *Year Book of the Royal Society* (1955) 198

(69a) GRUNDFEST, H.; *Arch. exp. Path. Pharmak.* 220 (1953) 136

(70) ECCLES, J. C., KATZ, B. and KUFFLER, S. W.; *J. Neurophysiol.* 4 (1941) 362

(71) KUFFLER, S. W.; *J. Neurophysiol.* 5 (1942) 18

(72) BROWN, G. L. and FELDBERG, W.; *J. Physiol.* 84 (1935) 12*P*

(73) BROWN, G. L. and MCINTOSH, G. C.; *J. Physiol.* 96 (1939) 10*P*

(74) HUTTER, O. F. and KOSTIAL, K.; *Biol. Bull.* 107 (1954) 292

(75) FATT, P. and KATZ, B.; *J. Physiol.* 118 (1952) 73

(76) NASTUK, W. L.; *Fed. Proc.* 13 (1954) 104

(77) CASTILLO, J. DEL and KATZ, B.; *J. Physiol.* 128 (1955) 396

(78) CASTILLO, J. DEL and KATZ, B.; *J. Physiol.* 126 (1954) 27*P*

(79) CASTILLO, J. DEL and ENGBAEK, L.; *J. Physiol.* 124 (1954) 370

(80) HUTTER, O. F. and KOSTIAL, K.; *J. Physiol.* 124 (1954) 234

(81) CASTILLO, J. DEL and STARK, L.; *J. Physiol.* 116 (1952) 507

(82) CASTILLO, J. DEL and KATZ, B.; *J. Physiol.* 124 (1954) 560

(83) NASTUK, W. L.; *J. cell. comp. Physiol.* 42 (1953) 249

(84) FATT, P. and KATZ, B.; *Acta physiol. scand.* 29 (1953) 117

(85) NASTUK, W. L. and ALEXANDER, J. T.; *J. Pharmacol.* 111 (1954) 302

(86) CASTILLO, J. DEL and KATZ, B.; *J. Physiol.* 124 (1954) 553

(87) CASTILLO, J. DEL and KATZ, B.; *J. Physiol.* 124 (1954) 574

(88) CASTILLO, J. DEL and KATZ, B.; *Nature*, Lond. 171 (1953) 1016

(89) JEFFRIES, G. H.; *J. Physiol.* 120 (1953) 611

(90) CASTILLO, J. DEL and KATZ, B.; *J. Physiol.* 123 (1954) 8*P*

(91) HODGKIN, A. L. and KATZ, B.; *J. Physiol.* 108 (1949) 37

(92) LLOYD, D. P. C.; *J. gen. Physiol.* 33 (1949) 147

(93) MAASKE, C. A., BOYD, T. E. and BROSNAN, J. J.; *J. Neurophysiol.* 1 (1938) 332

(94) BOYD, T. E., BROSNAN, J. J. and MAASKE, C. A.; *J. Neurophysiol.* 1 (1938) 497

(95) LUCO, J. V. and ROSENBLUETH, A.; *Amer. J. Physiol.* 126 (1939) 58

(96) FENG, T. P.; *Chin. J. Physiol.* 16 (1941) 341

(97) HUTTER, O. F.; *J. Physiol.* 118 (1952) 216

(98) LUNDBERG, A. and QUILISCH, H.; *Acta physiol. scand.* 30, suppl. 111 (1953) 111, 121

(99) COUTEAUX, R.; in *Le muscle—Etude de Biologie et de Pathologie* (1950) p. 173

(100) COWAN, S. L.; *J. Physiol.* 88 (1936) 3*P*
(101) FATT, P.; *J. Physiol.* 111 (1950) 408
(102) BURNS, B. D. and PATON, W. D. M.; *J. Physiol.* 115 (1951) 41
(103) NASTUK, W. L.; *Fed. Proc.* 10 (1951) 96
(104) BURGEN, A. S. V. and TERROUX, K. G.; *J. Physiol.* 120 (1953) 449
(105) KATZ, B.; *J. Neurophysiol.* 5 (1942) 169
(106) BARNES, T. C. and BEUTNER, R.; *Science* 110 (1949) 511
(107) FATT, P. and KATZ, B.; *Proc. Roy. Soc.* B. 140 (1952) 183
(108) BOYLE, P. J. and CONWAY, E. J.; *J. Physiol.* 100 (1941) 1
(109) CASTILLO, J. DEL and KATZ, B.; *J. Physiol.* 123 (1954) 70*P*
(110) CASTILLO, J. DEL and KATZ, B.; *J. Physiol.* 125 (1954) 546
(111) PERRY, W. L. M.; *J. Physiol.* 119 (1953) 439
(112) THESLEFF, S.; *Nature*, Lond. 175 (1955) 594
(113) PATON, W. D. M. and ZAIMIS, E. J.; *Pharmacol. Rev.* 4 (1952) 219
(114) PATON, W. D. M. and PERRY, W. L. M.; *J. Physiol.* 119 (1953) 43
(115) ZAIMIS, E. J.; *J. Physiol.* 122 (1953) 238
(116) FATT, P. and KATZ, B.; *J. Physiol.* 120 (1953) 171
(117) CASTILLO, J. DEL and KATZ, B.; (1955). In preparation
(118) FATT, P. and KATZ, B.; *J. Physiol.* 111 (1950) 46*P*
(119) HODGKIN, A. L.; *Proc. Roy. Soc.* B. 126 (1938) 87
(120) HODGKIN, A. L., HUXLEY, A. F. and KATZ, B.; *J. Physiol.* 116 (1952) 424

VOL
6
195

6

MODELS FOR BIOLOGICAL EXCITATION PROCESSES

U. F. Franck

CONTENTS

		Page
I.	GENERAL FEATURES OF BIOLOGICAL PROCESSES	172
II.	THERMODYNAMIC PROPERTIES OF OPEN DYNAMIC SYSTEMS	173
	(*a*) "Open systems"	173
	(*b*) Steady states in open dynamic systems	173
	(*c*) Stability of steady states	177
III.	SPECIAL NERVE MODELS	179
	1. Hydraulic models	180
	2. Electrical models	183
	3. Chemical models	184
	4. Electrochemical models	187
IV.	SPECIAL MODELS OF NERVE STRUCTURES	193
	(*a*) Saltatory propagation	194
	(*b*) Generation of high voltage in electric organs of electric fishes	198
	(*c*) Electrochemical models of rhythmical nerve activity	199
	(*d*) Models for interneural interaction and excitation transmission	203
	REFERENCES	204

6

MODELS FOR BIOLOGICAL EXCITATION PROCESSES

U. F. Franck

I. General Features of Biological Excitation Processes

A basic characteristic of living organisms is their ability to respond to stimuli emanating from their environment. This property in its elementary form, as shown by the single cell, is referred to as "excitability." It is well known that nerve cells are highly specialized in this respect. However, excitability may also be found in other cells, e.g., the muscle and gland cells, the electric cells of the electric fish, etc. There is a striking similarity in the excitability shown by all these cells, even in such cases as excitable plant cells where the organism apparently makes no use of this capacity. Therefore, it seems that excitability may be considered a very general characteristic of living cells.

Production of the excited state is subject to the All-or-Nothing Principle; i.e., a stimulus acting on the cell will bring about either the maximum excitation, or no excitation at all, depending on whether or not it exceeds a critical *threshold*. Accordingly, the stimulus causes only the release of the response and does not affect its intensity.

This behaviour is due to a specific instability of the cell membrane protoplasm towards certain (adequate) stimuli. In the inanimate world we are also acquainted with a great many so-called "unstable dynamic systems," in which the release of instability is strictly analogous to that in the plasma membrane. Apparently the latter, from the physico-chemical point of view, may be considered to be a similar system. This supposition is especially supported, since among the non-biological, unstable systems there exist a few which are able to reproduce functionally almost the entire excitation physiology, at least as far as the instability is concerned. The occurrence of such systems implies the possibility of constructing "nerve models" the study of which is likely to yield information about the nature of the physico-chemical mechanism of excitation processes.

In addition to the properties mentioned above, the nerve displays other typical characteristics of irritation phenomena. At first, these phenomena were thought to be independent properties of the nerve cell and were studied independently of each other. A better understanding, however, as demonstrated by the study of the models, proved

them to be consequences of the instability of the plasma membrane (K. F. BONHOEFFER[1–5], [40]). The most important of these are

(1) The threshold properties of the excitation release (validity of the All-or-Nothing Principle).

(2) The spontaneous, decrement-free propagation of the locally-released excitation state.

(3) The appearance of *refractory behaviour* (reduction of excitability as a result of the previous excitation).

(4) The *accommodation* phenomena (reduction of excitability as a result of a previous sub-threshold stimulus).

(5) The *rhythmical* behaviour (ability to respond under certain conditions to a steady stimulus with rhythmical recurring excitation).

Before entering upon a more detailed discussion of these phenomena, and their interpretation through models, a brief survey of the most important physico-chemical principles involved will be presented.

VOL. 6 1956

II. THERMODYNAMIC PROPERTIES OF OPEN DYNAMIC SYSTEMS

(a) *"Open systems"*

Seen from the thermodynamic standpoint, living organisms are to be considered as "open systems."[67] This expression is employed to denote a system which involves a steady irreversible flow of matter and energy causing a corresponding production of entropy. This also applies in full measure to the process in excitable cells discussed here. From a thermodynamic and reaction kinetic view, the problem of biological excitement, as well as all other questions involving energy change and metabolism, may be considered in the light of the thermodynamics of irreversible processes. Thus, a "nerve model" behaviour may be expected only from open entropy producing systems.

What is now meant by the above mentioned statement that excitable cells and their non-biological functioning models are "unstable dynamic systems"? This question may best be answered by the following concrete example:

(b) *Steady states in open dynamic systems*

If the conditions in an open system, macroscopically seen, are time independent, no real thermostatic equilibrium can exist, because irreversible processes take place continuously. Such a constant non-equilibrium state is referred to as a *steady state*. This steady state is always attained when the rates of all participating reactions occurring in the system remain constant. Both the steady state and true equilibrium possess the common feature in that each is able to display stability or instability. This will be illustrated by the following simple example.

As a model system we shall choose a container H (see Fig. 1) which is connected to the reservoir H_1, lying on a higher level, by means of the connecting tube R_1, and to the resevoir H_2, lying on a lower level, by means of the connecting tube R_2. The liquid levels of H_1 and H_2 as well as the temperature of the entire system are kept constant. Furthermore, the liquid shall possess a fixed viscosity, so that, the tube

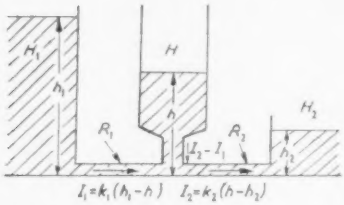


Fig. 1. Hydraulic model representing an open dynamic system.

diameters being sufficiently narrow, the Poiseuille Law may be considered valid. The flow intensities I in the tubes are then proportional to the difference in levels which exists between the reservoirs as shown

$$\text{for } R_1\text{: } I_1 = k_1(h_1 - h)\text{; for } R_2\text{: } I_2 = k_2(h - h_2) \qquad \ldots.(1)$$

where h_1, h_2, and h indicate the level heights in the corresponding reservoirs, and k_1 and k_2 the streaming conductance values (reciprocal streaming resistances) of the corresponding connecting tubes.

The container H, inclusive of the connecting tubes, represents our "open dynamic system." Because of its size the container possesses a certain storage capacity which, for a given net influx $I_1 - I_2$ causes the height of the level h to change with time according to

$$I_1 - I_2 = q \cdot \rho \cdot h \qquad \ldots.(2)$$

where ρ is the density of the liquid, and q the (constant) cross-section of the container H. Analogous to electric condensers we shall denote the product $\rho \cdot q$ as the capacity C of the system.

$$I_1 - I_2 = + C \cdot \dot{h} = (k_1h_1 + k_2h_2) - (k_1 + k_2) \cdot h \qquad \ldots.(3)$$

The friction of the liquid flowing in the connecting tubes causes a certain dissipation of energy Φ which, under the above made assumptions (validity of Poiseuille's Law, T = const.) can be directly calculated. This dissipation of energy Φ is equal to the product of the difference of the levels Δh and the flow intensity:

$$\begin{aligned} &\text{in } R_1\text{: } \Phi_1 = I_1(h_1 - h) = k_1(h_1 - h)^2 \\ &\text{in } R_2\text{: } \Phi_2 = I_2(h - h_2) = k_2(h - h_2)^2 \end{aligned} \qquad \ldots.(4)$$

Collectively, therefore:

$$\Phi = k_1(h_1 - h)^2 + k_2(h - h_2)^2$$

$$\Phi = (k_1 + k_2)\left(h - \frac{k_1h_1 + k_2h_2}{k_1 + k_2}\right)^2 + \frac{k_1k_2}{k_1 + k_2}(h_1 - h_2)^2 \quad \ldots.(5)$$

During the flow of liquid from H_1 to H_2 this amount will be yielded to the surroundings. For $T = \text{const.}$, this loss of energy by the system is related to the production of entropy by $S = \Phi/T$. The process

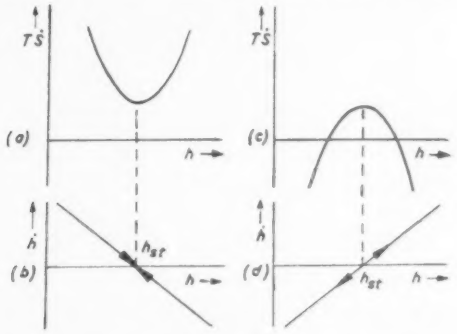


Fig. 2. Diagrams illustrating steady states in open systems. *a*, *b*—stable steady state; *c*, *d*—unstable steady state.

reaches a steady state when, in the container H, the influx I_1 equals the outflux I_2:

$$I_2 - I_1 = 0 \text{ and } k_1(h_1 - h_{\text{stat.}}) - k_2(h_{\text{stat.}} - h_2) = 0 \quad \ldots.(6)$$

$$\text{where } h_{\text{stat.}} = \frac{k_1h_1 + k_2h_2}{k_1 + k_2} \quad \ldots.(7)$$

Inserting eq. (7) into eq. (5) yields:

$$\Phi = T \cdot S = (k_1 + k_2)(h - h_{\text{stat.}})^2 + \frac{k_1k_2}{k_1 + k_2}(h_1 - h_2)^2 \quad \ldots.(8)$$

Therefore, the entropy production of our system is seen to be a function of the square of the level height in the storage container H. In a plot of $T \cdot S$ versus h (Fig. 2*a*) there results a parabolic curve (concave up), the coordinates of the minimum given by

$$h_{\text{stat.}} = \frac{k_1h_1 + k_2h_2}{k_1 + k_2};\ \Phi_{\text{stat.}} = T \cdot S_{\text{stat.}} = \frac{k_1k_2}{k_1 + k_2}(h_1 - h_2)^2 \quad \ldots.(9)$$

Therefore the steady state corresponds to the minimum entropy production in our system. A general derivation of this principle, which is important in its application to non-equilibrium processes, was first given by PRIGOGINE[8] employing the Onsager reciprocal relations.

If linear relations exist between the driving "forces" and the resulting "fluxes," as in the case of our model, the energy dissipation Φ of the system may be expressed in general form by the sum of the products of these forces and the corresponding fluxes[9]

$$\Phi = T \cdot S = \sum_i J_i X_i; \quad (X_i = \text{forces}; \ J_i = \text{corresponding fluxes}) \qquad \ldots.(10)$$

while between the fluxes and the forces the following "phenomenological relations" hold

$$J_i = \sum_{j=1}^{n} L_{ij} X_j \quad (i = 1, 2, \ldots, n) \qquad \ldots.(11)$$

where the L_{ij}'s are the so-called "phenomenological coefficients." Because of their dimensions they may be considered as conductances of the corresponding fluxes:

$$\text{dimensionally } [L_{ij}] = \frac{[J_i]}{[X_j]} \qquad \ldots.(12)$$

In our special model we may suitably choose as forces:

(1) the constant difference in height between the levels in the containers H_1 and H_2

$$X_1 = h_1 - h_2 \qquad \ldots.(13)$$

(2) the variable difference in height between the levels in the containers H and H_2

$$X_2 = h - h_2 \qquad \ldots.(14)$$

The corresponding fluxes are then:

$$J_1 = I_1 \text{ (flux from } H_1); \qquad J_2 = I_2 - I_1 \text{ (reservoir flow from } H) \qquad \ldots.(15)$$

Thus we have from (10)

$$\Phi = X_1 J_1 + X_2 J_2 = (h_1 - h_2) I_1 + (h - h_2)(I_2 - I_1) \qquad \ldots.(16)$$

The phenomenological coefficients L_{ij} are then

$$L_{11} = k_1 \qquad L_{12} = -k_1$$
$$L_{21} = -k_1 \qquad L_{22} = (k_1 + k_2) \qquad \ldots.(17)$$

According to ONSAGER's reciprocal relations the following is valid

$$L_{12} = L_{21} = -k_1$$

Combining equations (17), (11), and (10) we obtain the energy dissipation Φ and the entropy production $\dot{S}$

$$\Phi = T \cdot S$$

$$= L_{11}X_1^2 + (L_{12} + L_{21})X_1X_2 + L_{22}X_2^2 \qquad \ldots.(18a)$$

$$= k_1(h_1 - h_2)^2 - 2k_1(h_1 - h_2)(h - h_2) + (k_1 + k_2)(h - h_2)^2 \ldots.(18b)$$

$$= (k_1 + k_2)\left(h - \frac{k_1h_1 + k_2h_2}{k_1 + k_2}\right)^2 + \frac{k_1k_2}{k_1 + k_2}(h_1 - h_2)^2 \qquad \ldots.(18c)$$

This expression is identical with (5).

(c) *Stability of steady states*

Since the factor $(k_1 + k_2)$ in the term involving the second power of h is a conductance, it is normally positive. It may be readily shown that the steady state corresponding to minimum entropy production is stable; for, after each change forced upon it by an external transient influence, the system, when left to itself, will return to its original state. In our hydrodynamic model there results, e.g., when $h > h_{\text{stat.}}$, both a decrease of influx to, and an increase of outflux from H, so that the level in H falls until $h_{\text{stat.}}$ is again attained. The reverse occurs for the case $h < h_{\text{stat.}}$, the level returning again to $h_{\text{stat.}}$. This relation appears directly from equation (3)

$$Ch = (k_1h_1 + k_2h_2) - (k_1 + k_2) \cdot h \qquad \ldots.(19a)$$

$$h = \frac{k_1 + k_2}{C}\left(\frac{k_1h_1 + k_2h_2}{k_1 + k_2} - h\right) \qquad \ldots.(19b)$$

where $\dot{h} = \dfrac{1}{\tau}(h_{\text{stat.}} - h)$

$$\text{and } \tau = \frac{C}{k_1 + k_2}\left(\text{time constant}\right) \qquad \ldots.(19c)$$

Therefore the change of h with time depends on h itself. In Fig. 2*b* $\dot{h}$ is plotted against h as given by eq. (19). The plot yields a straight line with negative slope intersecting the h axis at $h_{\text{stat.}}$. Therefore for values of h greater than $h_{\text{stat.}}$, $\dot{h}$ is negative. Thus, there occurs a decrease of h and with it a return to $h_{\text{stat.}}$. For values of h smaller than $h_{\text{stat.}}$, $\dot{h}$ is positive, causing an increase in h which returns to $h_{\text{stat.}}$. Therefore we may write for the stability condition:

$$\frac{d\dot{h}}{dh} < 0 \text{ for } \dot{h}(h) = 0 \qquad \ldots.(20)$$

On the other hand, unstable steady states will occur when a curve, such as that in Fig. 2*d*, displays a positive slope at the point where it

intersects the h-axis. Therefore, we have the following condition for instability

$$\frac{dh}{dh} > 0 \text{ for } \dot{h}(h) = 0 \qquad \ldots.(21)$$

For the special case of our model this condition is fulfilled when the following is valid

$$(k_1 + k_2) < 0 \text{ or } k_2 < -k_1 \qquad \ldots.(22)$$

This means that a system involving unstable steady states must include a negative conductance. If we assume, e.g., that in our model system k_1 is negative ($k_1 = -k_1^*$) then equations (5) and (18) will take the following form:

$$\Phi = T \cdot S = -(k_1^* - k_2)(h - h_{\text{stat.}})^2 + \frac{k_1^* k_2}{k_1^* - k_2}(h_1 - h_2)^2 \qquad \ldots.(23)$$

Furthermore, if we assume that the conductances k_1 and k_2 in the vicinity of

$$h_{\text{stat.}} = \frac{k_1^* h_1 - k_2 h_2}{k_1^* - k_2}$$

may be regarded as constant, and also that the Poiseuille law is fulfilled, a plot of $T \cdot S$ versus h (Fig. 2c) will exhibit the parabolic form (concave down) as seen. In this case the unstable steady state corresponds to the curve maximum, i.e., the maximum entropy production.[10]*

At first sight, negative conductances and resistances appear incapable of realization, for their existence must result in a continuous decrease of entropy. The latter, of course, contradicts the Second Law of thermodynamics. For this reason negative conductances are impossible in closed systems. In open systems, however, the net entropy production may also be negative without contradicting the Second Law. In these cases PRIGOGINE's "generalized entropy function" may be applied:[11]

$$dS = d_e S + d_i S \qquad \ldots.(24)$$

Here, $d_e S$ denotes the entropy changes brought about by transport, and $d_i S$ the entropy changes caused by irreversible processes within the system. According to the Second Law, $d_i S$ is always greater than zero. On the other hand, $d_e S$ may be greater than, equal to, or less than zero. The latter value will prevail when substances of high free energy content are flowing into the system. We recognize, therefore, that negative conductances are realizable in open systems.

By the expression "unstable dynamic system," employed at the outset to characterize excitable biological objects and their non-biological models, we mean open systems of the type described above.

* The statements of equations (8) and (23) that the stability or instability of steady states are characterized by the minima or maxima of the entropic production is in general held only for open systems with linear (positive or negative) conductance.

These systems, owing to negative conductances, are able to assume unstable steady states corresponding to eq. (21).

III. Special Nerve Models

In the case of nerves, Hodgkin, Huxley and Katz[12] have determined the cation flux (I^+) through the nerve membrane as a function of the membrane potential, by means of measurements on the giant axon of the squid, *Loligo forbesi*. They obtained the curve shown in Fig. 3.

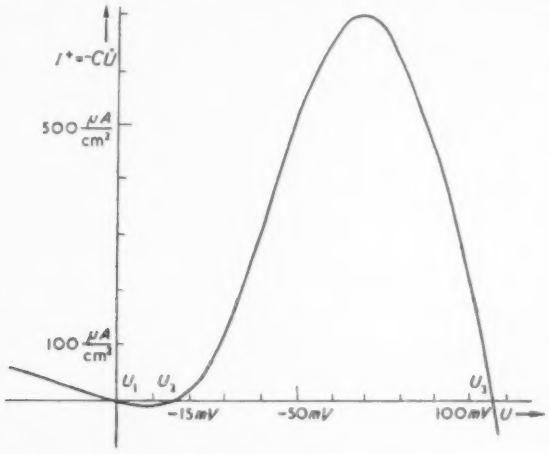


Fig. 3. Relation between U and $\dot{U}$ observed on the giant axon plasma membrane of *Loligo forbesi* (Hodgkin, Huxley and Katz, *Arch. Sci. physiol.* 3 (1949) 136). The diagram given in the original paper as Fig. 6 was re-plotted so as to correspond to Fig. 5. Abscissa: membrane potential; ordinate: ion flux $I^+ = -C_m \cdot \dot{U}$

Since the cations (in this case Na^+ and K^+), migrating through the membrane under the influence of an applied potential difference, carry electric charges, they will discharge the membrane double layer capacity $C_m (\sim 0{\cdot}96\ \mu F/cm^2)$ according to eq. (2):

$$I^+ = -C_m \dot{U} = f(U) \qquad \ldots . (25)$$

Therefore, the experimental curve yields a relationship between the variable U and its time derivative corresponding to eq. (3) and Fig. 2. We see that the curve intersects the U-axis at three points, and therefore possesses three steady states ($U_1 = 0$, $U_2 = -15$ mV, $U_3 = -106$ mV). For U_1 and U_3, we observe that the stability condition (20) is fulfilled since, at these points, $d\dot{U}/dU < 0$. On the other hand U_2 is unstable, for here, according to (21), $d\dot{U}/dU > 0$. Therefore, at -15 mV the nerve membrane actually behaves as if it has a negative conductance: $G_m \equiv dI^+/dU < 0$.

We find the same situation, generally, in the case of all systems proposed as nerve models and suitable for this purpose. We can most easily obtain a general understanding of these systems if we consider the hydraulic model employed in the previous chapter. All regularities demonstrated in this case can be carried over, in principle, to all analogously constructed, open systems. The hydrostatic pressure and hydrodynamic flux could, e.g., be replaced by the "forces" and "fluxes" compiled in the following table:

TABLE I

Examples of Conjugated Forces and Fluxes in Open Systems

Type of System	*"Force"*	*"Flux"*	*Conductance*	*Negative conductance may occur through:*
Mechanical	Force (pressure)	Mass flux	Gliding (= reciprocal friction)	Negative friction
Electrical (*a*) Capacitive	Voltage	Electrical current	Conductance	Dynatron behaviour
(*b*) Inductive	Electric current	Voltage	Resistance	Arc behaviour
Thermal	Temperature	Heat flow	Thermal conductivity	
Reaction kinetic	Affinity	Reaction rate	Reaction coefficient	Autocatalysis
Diffusion, membrane permeability	Concentration	Rate of diffusion	Diffusion coefficient	Excitable membrane

The last column of the table reveals the possible types of negative conductances which, under the discussed conditions, can lead to unstable steady states. The table illustrates the great variety of nerve models that can be constructed.

1 *Hydraulic models.*

These models have the advantage of clearly demonstrating the connection between the operating "forces" and the resulting "fluxes" so that they can be employed in the case of hypothetical experiments—for the functional interpretation of physiological processes (A. V. HILL[13] and H. GERSTNER,[14] among others). Negative conductances can easily be realized in these cases by means of suitable overflow siphons or float

valves (compare Fig. 4). By use of such devices a dependence of one of the flow conductances on the hydrostatic pressure can be achieved.

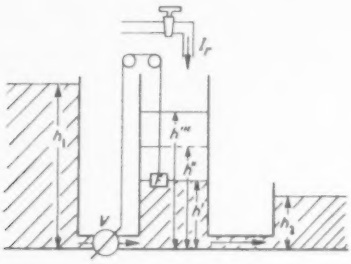


Fig. 4. Hydraulic model representing an unstable open system. The float F is coupled with the valve V in a way effecting a relationship between k_1 and h which, as in eq. (26), involves the second power of h.

Let us assume, e.g., that the following relation exists between k_1 and h

$$k_1(h) = \alpha + \beta h + \gamma h^2 \qquad \ldots.(26)$$

Then, according to (19), the expression for $\dot{h}(h)$ will involve h^3

$$C \cdot \dot{h} = -\gamma h^3 - (\beta - \gamma h_1)h^2 - (\alpha - \beta h_1 + k_2)h + \alpha h_1 + k_2 h_2 \qquad \ldots.(27)$$

the three roots of which correspond to steady states. Furthermore, let us suppose that these are real and that α is positive; then we obtain in the corresponding $\dot{h}$ versus h plot a curve (Fig. 5*a*) analogous to that of Fig. 3. Of the three points of intersection with the h-axis both h' and h''', according to condition (20), are proved to be stable steady states. On the other hand, at h'' the system as seen from (21) is unstable.

The adequate "stimulus" of such a system, the resting state of which is in h', consists in an elevation of the level h through an external disturbance (influx of a certain quantity of fluid Q in the system container H). Such a stimulus is "supra-threshold" when the resulting increase in level surpasses the unstable state h''. After discontinuation of the stimulating action the system then finds itself in a state with a positive value of $\dot{h}$. It, now no longer returns to h', but passes over into the "excited state" h'''. On the other hand, in the case of sub-threshold stimulation, h'' is not attained. Because $\dot{h} < 0$, the system returns to h'.

Under the action of a constant stimulating current I_r the relationship $\dot{h} = -\frac{1}{C}(I_2 - I_1 - I_r)$ corresponding to (3) is valid. Consequently, such a current, in the $\dot{h}$ versus h plot, causes an upward parallel displacement of the curve by the amount $\dot{h} = \frac{I_r}{C}$. Thus h'

and h''' are moved to the right while h'' is displaced to the left. For a certain stimulating current, h' and h'' will coincide (curve 2, Fig. 5*b*), while further increase of this current will cause them to disappear (positive values of $\dot{h}$ to h''' being assumed everywhere (curve 3.)) This minimum stimulating current which, when flowing for an infinite length of time, causes the excitation of the system, is referred to as "rheobase"

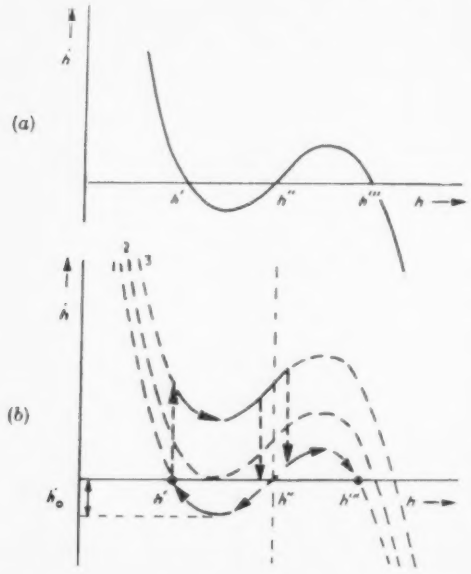


Fig. 5. (*a*) Relation between $\dot{h}$ and h in the model outlined in Fig. 4. (*b*) Paths of state for sub-threshold and supra-threshold stimuli. Dislocation of the basic curve (1) by $\dot{h}_0$ to (2) corresponds to the effect of a rheobasic stimulating current, dislocation to (3) to the effect of a supra-rheobasic stimulating current ($I_\tau = C \cdot \dot{h}$).

in the neurophysiology of excitation. Therefore, in our model, this is directly obtained by multiplication of the quantity $\dot{h}_0$—indicated in Fig. 5*b*—by the capacity C of the system:

$$\text{rheobase: } I_\tau^0 = -C \cdot \dot{h}_0 \qquad \ldots.(28)$$

Each supra-threshold stimulating current $I_\tau > I_\tau^0$ must last for a certain stimulating time τ_r if it is to cause excitation with a resulting elevation of the liquid level from h' to h''. The quantity of fluid $Q_r = I_\tau \cdot \tau_r$ flowing in during this time, will be denoted as the "threshold," and will be the lower, the more rapidly it is supplied to the system. In our model, the limiting value ("minimum threshold")

attained for $\tau_r \to 0$ is given by the quantity of liquid Q_r^0 lying between h' and h''

$$\text{minimum threshold: } Q_r^0 = (h'' - h')\rho q = (h'' - h') \cdot C \quad \ldots . (29)$$

In the h versus h plot the excitation process can be clearly specified by paths of state. This is drawn in Fig. 5*b* for sub-threshold and supra-threshold cases.

For short stimulating times and corresponding high excitation currents, the $i \cdot t$ law is valid as long as the changes caused by the fluxes I_1 and I_2 during the time τ_r can be neglected

$$I_r \cdot \tau_r = Q_r^0 = \text{const.} \quad (\tau_r \to 0) \qquad \ldots . (30)$$

Combining such hydraulic models, by means of appropriate siphons, the propagation of excitation process can also be demonstrated without difficulty. The model, however, in the simple form sketched here, does

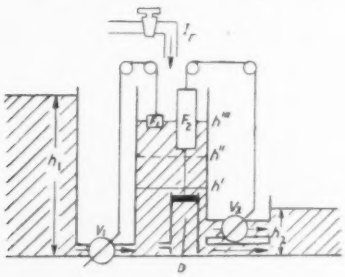


Fig. 6. Hydraulic model representing an unstable open system with "recovery." The float F_2 and the valve V_2 are coupled by a delayed action mechanism (D) in a way as to achieve, after every "excitation," the return of the level to h'.

not possess repeated excitability, refractory behaviour, and accommodation. Through the incorporation of a suitable discharge valve (D, Fig. 6), operated with delayed action by a float mechanism depending on $(h - h')$, the model can be completed in this respect. Hydraulic models of this kind were employed by A. V. HILL[13] to illustrate the theory of accommodation developed by him. Fig. 6 schematically displays a nerve model complete in the sense of the Hill theory.

2. *Electrical models*

Under certain conditions negative electric conductances (or resistances) are known to arise in vacuum tubes (e.g., dynatron, negatron, and negadyne circuits) and also in electric arcs and gas discharge tubes.[15–19] These possess non-monotonic current voltage curves, namely in the case of maximum or minimum values either for I or U. For the former N-shaped "dynatron" characteristics result while for the latter S-shaped "electric arc" characteristics are obtained (Fig. 7).

Therefore, according to this notation, the nerve possesses (from Fig. 3) a dynatron characteristic.

In principle, the considerations advanced in the case of the hydraulic model can also be applied to electric circuits with negative conductances. The systematic scheme shown in Fig. 1 is then replaced by

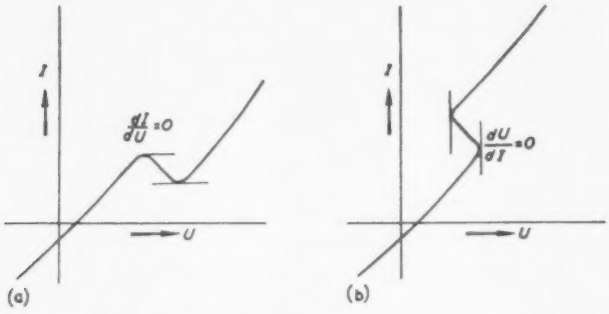


Fig. 7. (*a*) Dynatron characteristic; (*b*) Electric arc characteristic.

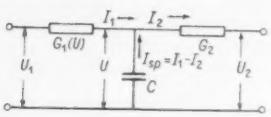
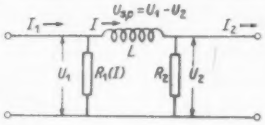
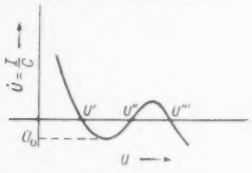
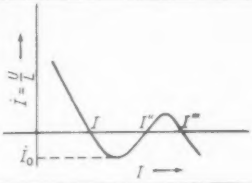
equivalent fourpole circuit diagrams (Table II). Since dynatron type conductances are voltage dependent they must be connected with capacities whose charge condition is also voltage dependent. Consequently, the current dependent electric arc resistances must be combined with inductances. Between both types of connections the well-known reciprocity between current and voltage, or resistance and conductance, respectively, exists. Considering these relations we obtain for these variables completely analogous formulæ which we derived for the hydraulic models. If we assume here that the voltage U depends upon the conductance G_1 and that the resistance R_1 upon the current I, then we obtain the expressions cited in Table II which are completely analogous to the equations (25), (27), (28), (29).

Such non-linear unstable fourpoles have been used at different times for investigations of excitation physiology models. In this manner, VAN DER POL and VAN DER MARK[20] very clearly demonstrated, with neon lamp circuits in the form of a heart model, the idea that biological rhythmical phenomena are essentially relaxation oscillations. A. BETHE[21] thoroughly studied such neon lamp arrangements as models for biological irritation and rhythmical behaviour. In recent times, by linking together neon lamp circuits, GREY WALTER[22] has constructed a nerve model which is able to reproduce almost all electric excitation phenomena of the nerve (Fig. 8).

3. *Chemical models*

Analogous to the negative conductances of the unstable hydraulic and electric systems, there occur in reaction kinetics, in the case of

TABLE II

Dynatron conductance	*Arc resistance*
	
	
$I_{sp} = I_1 - I_2 = C\dot{U}$ $= (G_1 + G_2)U - G_1U_1 - G_2U_2$	$U_{sp} = U_1 - U_2 = L\dot{I}$ $= (R_1 + R_2)I - R_1I_1 - R_2I_2$
$G_1(U) = \alpha' + \beta'U + \gamma'U^2$ $CU = -\gamma'U^3 - (\beta' - \gamma'U_1)U^2$ $- (\alpha' - \beta'U_1 + G_2)U + \alpha'U_1$ $+ G_2U_2$	$R_1(I) = \alpha'' + \beta''I + \gamma''I^2$ $L\dot{I} = -\gamma''I^3 - (\beta'' - \gamma''I_1)I^2$ $- (\alpha'' - \beta''I_1 + R_2)I + \alpha''I_1$ $+ R_2I_2$
rheobase: $I_r^0 = -C \cdot \dot{U}_0$ minimum threshold: $Q_r^0 = C(U'' - U')$ (amp-sec.)	rheobase: $U_r^0 = -L \cdot \dot{I}_0$ minimum threshold: $Q_r^0 = L(I'' - I')$ (volt-sec.)

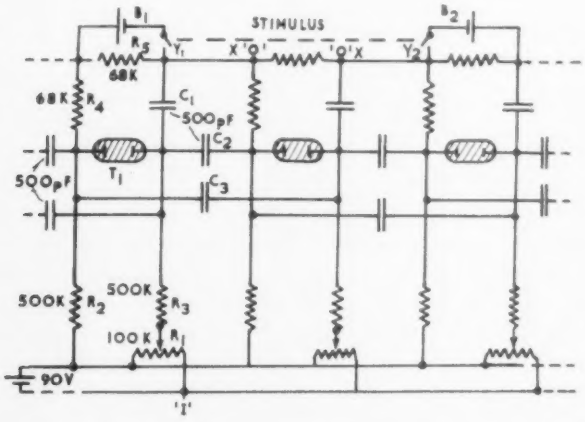


Fig. 8. Electric nerve model designed by GREY WALTER. The figure shows only three sections of the large number of neon lamp circuits forming the device. (GREY WALTER, *The Living Brain*, London.)

auto-catalytic reactions, negative reaction coefficients (coefficients d'entraînement, PRIGOGINE[11]). Such negative rate constants are found in the case of reactions catalysed by the reaction products themselves, and also in the case of exothermic reactions proceeding non-isothermally, where self-commencement is brought about when the resulting heat of reaction is not adequately conducted to the surroundings. Through the investigation of the heat balance between evolved and conducted heat VAN'T HOFF, for the first time, properly explained the occurrence of a thermal explosion as the exceeding of an unstable state in an open, exothermic reaction system.[23–25] For this type of autocatalytic

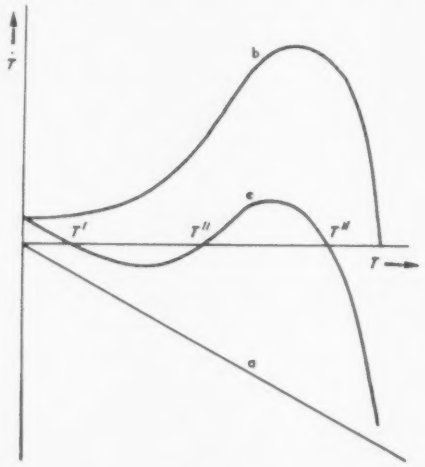


Fig. 9. Relation between T and $\dot{T}$ in thermal explosion.

reaction the conditions appear especially clear. These can be easily illustrated in a graphic manner so that we shall discuss them briefly here. For simplicity we shall assume that the specific heat of the reaction remains always constant and that the loss of heat through conduction to the surroundings is proportional to the temperature difference existing between the inside and outside of the system. We can then picture the heat efflux in a $\dot{T}$ versus T plot (Fig. 9) as a straight line with a negative slope (curve a). On the other hand, in the case of exothermic reactions the heat evolution increases with temperature generally in a non-linear manner (curve b). We obtain the total behaviour of the system by summation of both curves (curve c). The three points of intersection are steady states, of which the second is unstable while the first and the third are stable. The unstable state corresponds to the ignition temperature. Explosible gas mixtures obey the All-or-Nothing-Principle[1] and display the phenomenon of decrement-free transmission in the form of a self-propagating explosion

zone. In particular, when a fresh supply of gas is provided in a suitable manner, the system acquires the property of recovery and under certain circumstances, rhythmical response.

A chemical reaction behaving in a similar way—the oxidation of formaldehyde by concentrated nitric acid—has been extensively studied by G. VERWORN[26] with respect to the properties analogous to those of nerve excitation. The adequate stimulus will be attained here by the dropwise addition of nitrous acid which, as is known, markedly catalyses the oxidation effect of nitric acid so that the reaction commences at the point of addition. But by means of the latter, fresh nitrous acid is formed, which in turn, diffuses into the not yet affected vicinity and there initiates the reaction until through such successive processes, the total reaction mixture has reacted.

A few years earlier LUTHER[27] had already experimented with such reaction mixtures, including permanganate-oxalic acid solutions, which also displayed propagation phenomena. In this case the resulting manganous ion catalyses the reduction of the permanganate ion in a manner corresponding to that in which the nitrous acid catalyses the nitric acid reduction. In both examples the diffusion of the resulting catalyst assumes the same rôle which the heat of conduction plays in the case of the propagation of thermal explosions, so that the former functions also under isothermal conditions.

VOL. 6 956

4. *Electrochemical models*

In electrolytic processes non-metallic, semi-conducting surface layers are often deposited on electrodes. These layers are able to

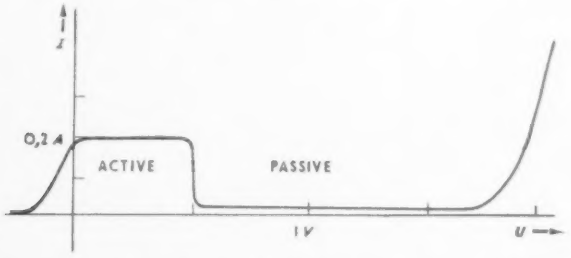


Fig. 10. Current voltage characteristic of iron in 1 *N* sulphuric acid, a dynatron type characteristic.

influence the electrode behaviour so that dynatron type current-voltage characteristics are realized. This is especially true in the case of metals capable of being rendered passive, e.g., iron, nickel, cobalt, gold, etc.,[28] upon which electron conducting, scarcely soluble oxide layers are formed at certain "passivation potentials." As a typical example, Fig. 10 shows the steady current-voltage characteristic of

iron in sulphuric acid which possesses, at the passivation potential $E_h = 460$ mV, a negative slope.[28], [29] A polarization circuit into which such a dynatron type electrode is introduced (Fig. 11*a*) corresponds to

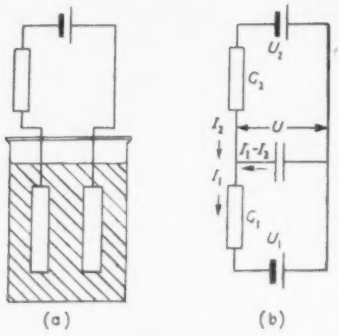


Fig. 11. Electrolytic circuit (*a*) and its equivalent circuit (*b*).

the equivalent circuit indicated in Fig. 11*b*. In principle this has the same appearance as the scheme shown in Table II for voltage dependent, electric systems. Further, in Fig. 12*a*, the partial currents I_1, I_2, and

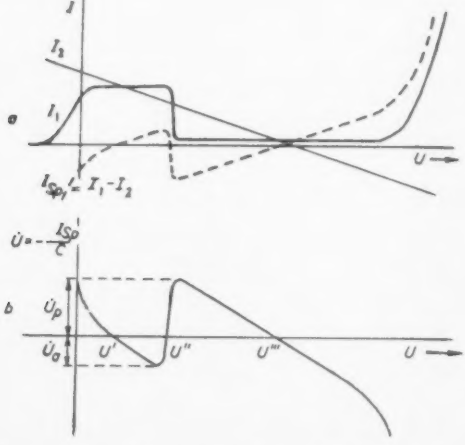


Fig. 12. Diagram illustrating instabilities of electrodes.

$I_{sp} = I_1 - I_2$, flowing in the electrode circuit, are plotted versus the potential U under the assumption that G_1 possesses a current-voltage characteristic corresponding to that of the iron electrode. Since I_{sp} is now equal to $C \cdot \dot{U}$ the dynamic characteristic of the electrode circuit in the U versus U plot can be directly constructed.[10] Analogous to Figs. 2 and 3 we also obtain here a non-monotonic curve with one

unstable and two stable steady states. Therefore the electrolytic circuits with dynatron type electrodes are also unstable systems and, as we shall see, have proved themselves exceptionally well suited for utilization as models for biological excitation processes.

Anodic or cathodic currents appear as stimuli depending on which of the two stable states U' or U''' is selected to represent the resting state of the nerve. The dynamic characteristics will be raised in a parallel fashion through anodic stimuli and lowered in a similar way by the application of cathodic stimuli.[2] In the first case (resting state in U') the rheobase is given by $I_r^0 = \dot{U}_a \cdot C$; in the second (resting state in U''') it is given by $I_r^0 = \dot{U}_p \cdot C$ (Fig. 12*b*). In the case of models employing passive iron electrodes one generally chooses the passive condition (U''') as being analogous to the unexcited nerve. Thus, the excitation of the latter corresponds to the activation of the passive iron through cathodic stimuli.

Already in 1827, G. WETZLER[30] had observed that the activation of a passive iron wire immersed in concentrated nitric acid, when initiated at some arbitrary position by contact with a more reactive metal, was spontaneously propagated over the entire metal surface. But the fundamental suggestion that this propagation process might provide a model for propagation in nerves was advanced much later, and for the first time, by W. OSTWALD, in 1900.[31-34]

Later, the Ostwald nerve model was again taken up by R. LILLIE[35], [33], who investigated thoroughly its physiological analogous properties. Surprised, he found that it was able to reproduce completely the biological excitation processes. However, it was only recently that K. F. BONHOEFFER explained the true physico-chemical reasons for this remarkable similarity.[1-5] He recognized the kinetic instability as common to both systems and showed that most of the excitation phenomena are a direct consequence of this instability, in the sense as already presented in Section I.

About 1870, L. HERMANN[37] had already developed a theory for the propagation of excitation in nerves which, in outline, corresponds to modern views. From evidence already known at that time—that each excited position in the nerve assumed a more negative potential than a non-excited position—HERMANN suggested the presence of currents flowing between these regions. Such currents would then, in turn, operate in an exciting manner on their non-excited surroundings so that a successive displacement of the excitation front is brought about. For the passive iron wire the relations are quite similar. Here, the active positions possess a more negative potential than the passive so that, in the case of partially activated iron surfaces, local currents flow between the active and passive regions. For the latter these are cathodic. Therefore they operate there in an activating manner so

that a pattern of the same kind as in the Hermann "Strömchentheorie" results (Fig. 13*a*).

Seen from an electrochemical standpoint, the excitable cell as well as the Ostwald model consists of two electrically conducting phases which are separated by a thin intermediary layer, or membrane, which is unstable against electric stimuli (Fig. 13*b*). In the case of nerve it is the invisible, thin protoplasm membrane; for passive iron, it is likewise an invisible, thin (thickness $\sim$ 50 Å) oxide layer. As we

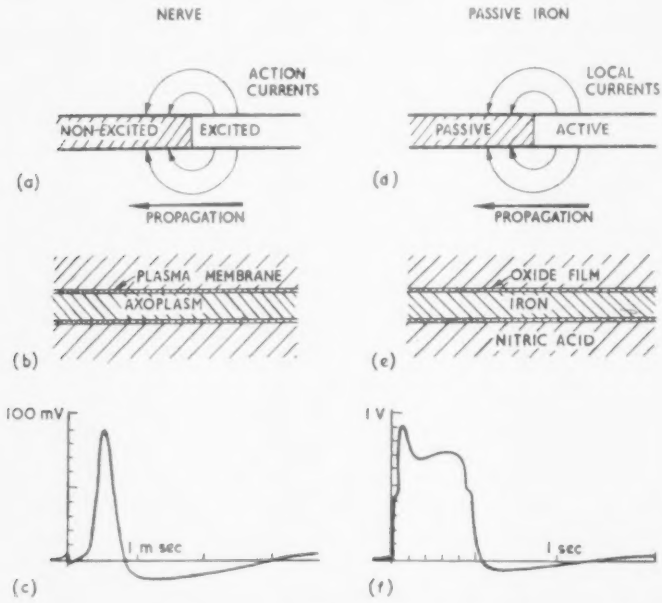


Fig. 13. Comparison between nerve and OSTWALD's iron wire model: *a*, *d*—analogy in local currents; *b*, *e*—structural analogy; *c*, *f*—analogy in the form of the "action potential."

have seen, the iron model and the nerve, beyond the instability, display agreement in electrochemical and structural respects.[34] Obviously, here lies one of the reasons for the extraordinary analogous properties found.

If one records the potential during the activation of iron there results an activation curve[38] similar to the action potential of nerve[39] (Fig. 13*c*, *f*). As in the case of nerves, the resting potential U_0 is attained only after passing through a positive "after potential" rather than immediately in the descending phase. Since the relationship between the nerve and the wire model only depends on a formal analogy, the time and potential values corresponding to one another in both systems are different. Table III, which was compiled from

data of BONHOEFFER,[40] HODGKIN, HUXLEY, KATZ,[39] and MURALT,[41] presents a comparison of the most important values:

TABLE III

Comparison of Characteristic Data for Nerve and the Ostwald Model

	Nerve Axon of Loligo	*Ostwald Model* Iron in 14 *N* HNO_3
Spike potential (with reference to the resting potential	80–100 mV	980–1000 mV
"Ascending phase"	0·1 msec.	~ 10 msec.
"Descending phase"	1–2 msec.	0·5 sec.
Rheobase	$3{\cdot}5 \cdot 10^{-5}$ amp/cm²	~ $5 \cdot 10^{-3}$ amp/cm²
Minimum threshold	$2 \cdot 10^{-8}$ coulomb/cm²	$2 \cdot 10^{-4}$ coulomb/cm²
Refractory time	$5 \cdot 10^{-3}$ sec.	10^2–10^3 sec.
Unstable potential (with reference to the resting potential)	− 15 mV	− 500 mV

As we see from Fig. 13*f*, iron possesses the capacity, after having been activated, to return again to the passive state. Investigations by K. F. BONHOEFFER and co-workers,[42–45] and K. VETTER[46–48] have shown that the nitrous acid plays a decisive rôle in the repassivation process. It was already shown, in connection with the model of VERWORN, that the oxidation effect of nitric acid was markedly increased in the presence of nitrous acid. In the active phase, while the nitric acid attacks the iron in rather vigorous fashion, nitrous acid forms in increasing amounts on the iron surface and autocatalytically accelerates the oxidation process of the iron so that finally passivation again occurs. The resulting increased oxidation strength decays afterwards in a slow manner: this is borne out by the fact that freshly passivated iron is more difficult to activate than that in a recovered condition. Consequently, after each activation the threshold and rheobase are raised as observed in the refractory nerve. Production of nitrous acid and threshold elevation respectively, also take place at sub-threshold stimuli, so that in the case of the Ostwald model accommodation as well as refractory behaviour is found[44].

In Fig. 13*d* the indicated local current mechanism for the transmission of activation concerns only the propagation of the leading part of the activation front in the activity zone. However, in the case of the subsequent passivation wave, only the repassivation process, and not the local current mechanism, is involved. Therefore the length of the developed zone is determined by the velocity with which the activation front moves forward and by the time of repassivation. Consequently, activity zones moving quickly are longer than those

moving slowly,[49–51] provided that the repassivation time is the same. By means of two reference electrodes placed at the wire, with a definite distance from each other, the propagation process can be followed on an oscillograph and the velocity measured. Fig. 14 shows two typical examples: (*a*) that of the Ostwald model,[49] (*b*) that of a cobalt wire

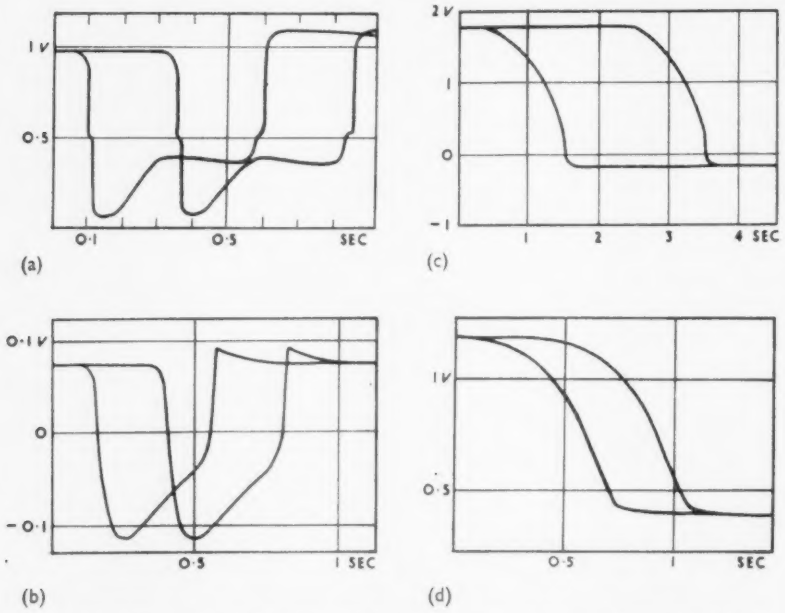


Fig. 14. Propagation of active zones on passive metal electrodes: (*a*) on iron in 12 *N* HNO_3; (*b*) on cobalt in 1 *N* HCl + 1·2 *N* CrO_3; (*c*) on iron polarized anodically in 1 *N* H_2SO_4; (*d*) on gold polarized anodically in 1 *N* HCl + 2 *N* NaCl. The systems (*a*) and (*b*) involve recovery; the systems (*c*) and (*d*) do not.

immersed in a mixture of chromic and hydrochloric acids.[52] Since the activity zones successively pass the reference positions at different times, the corresponding recorded activation curves are displaced in a parallel fashion.

There are also electrochemical systems that exhibit no repassivation and therefore no accommodation and refractory behaviour. On account of their greater simplicity these offer essential experimental advantages for the study of activation and propagation processes by local currents.[53] In Table VI examples of suitable model systems, with and without repassivation, are given.[55]

Through the suitable combination of electrodes and electrolytes or polarization currents, respectively, one has a certain freedom in the choice of properties which the model system shall possess. For example,

TABLE IV

Examples of Electrochemical Model Systems

Metal	*Electrolyte*	*Condition after activation*
Iron[49]	conc. HNO_3	Repassivation
Iron[55]	1 *N* CrO_3	Repassivation
Iron[53]	H_2SO_4	No repassivation
Cobalt[52]	1 . 2 *N* CrO_3 + 1 *N* HCl	Repassivation
Gold[54]	1 *N* HCl + 2 *N* NaCl	No repassivation
Zinc[56]	2 *N* NaOH	No repassivation

model systems can be devised obeying the different strength-latency relation formulæ which have been proposed for the behaviour of nerve. The relation $i = a + \frac{b}{t}$ (i = stimulus current, t = stimulus time, a and b are constants) given by WEISS[57] or HOORWEG,[58] respectively, will be fulfilled by a great number of electrode systems, e.g. iron in sulphuric acid at anodic passivation[59] (see Table IV)

$$\tau_p = \frac{b}{I_p - I_0} \quad \text{or} \quad I_p = I_0 + \frac{b}{\tau_p}, \text{ respectively,}$$

where τ_p is the passivation time, I_p the passivation current density, and I_0 and b are constants. On the other hand, the cathodic activation of this system[59] obeys BLAIR's law:

$$\tau_a = \frac{1}{k} \ln \frac{I_a}{I_a - Rh}$$

(τ_a = activation time, I_a = cathodic current, Rh = rheobase, k = constant). The behaviour of OSTWALD's model is more complex. In the first approximation it is described by HILL's law[13] which is known to include recovery.

IV. Special Models of Nerve Structures

It appears that the basic excitation reaction of the protoplasm also plays the essential functional rôle in the case of complicated neural nets. Since properly suitable models are available in electro chemical unstable systems, it should thus be possible to reproduce such structures. With regard to the electrical properties of the tissue concerned, we can insert everywhere in the model, where excitable membranes are suspected, passive iron electrodes in contact with concentrated nitric acid, and compare the entire system in its function with the biological structure. A few specific examples designed in this way are given in the following.

(a) *Saltatory propagation*

The fast conducting peripheral nerves of vertebrates, i.e., the so-called "myelinated" nerves, possess, as is known, a periodic rather than a uniformly continuous structure.[41], [61] As a result, a comparison with an iron wire is not directly possible. In Fig. 15 this structure, markedly

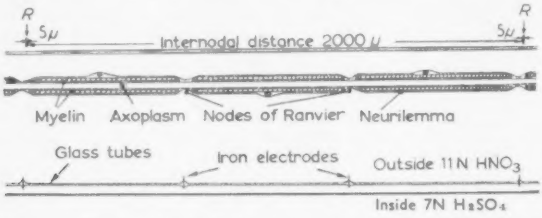


Fig. 15. Diagram illustrating saltatory propagation. (*a*) myelinated nerve, schematically; (*b*) myelinated nerve, giving the correct size of the nodes relative to the internodal distance; (*c*) glass tube model.

simplified, is schematically represented. Essentially such a nerve consists of a continuous membrane tube (axon) which is enveloped by the myelin sheath. Constrictions, referred to as the nodes of Ranvier, occur in this sheathing at regular intervals. At these points the myelin sheath is broken so that here, and only here, is the nerve fibre capable of being electrically excited, since myelin is a rather good insulator. There is a great deal of experimental evidence that the propagation of excitation at such nerves does not take place continuously, but rather in a saltatory fashion from node to node. Already in 1925 LILLIE[62] expressed this supposition when he observed, in his experiments with the Ostwald model, that the propagation of activation took place in this manner if he covered the iron wire, in sections, with glass tubing. It is especially interesting to note that through this partial insulation of the wire surface by glass tubing the velocity of the propagation was not lessened but clearly increased.

According to this idea, the electrical aspect of the myelinated nerve consists of an electrolytic conductor (Axoplasm) which is separated by an insulating sheath from a conducting outer medium. At equal intervals both electrolyte chambers are connected with one another through excitable membranes. In this way we can construct a model[63] by replacing the myelin hollow cylinder by a glass tube, in the walls of which small iron electrodes are sealed at regular intervals, joining electrically the tube interior with the outer electrolyte. The tube will be filled with some arbitrary, electrolytically conducting fluid and immersed fully in concentrated nitric acid. The iron electrodes are then passivated when brought into contact with the nitric acid and function in this form as models of the excitable membranes of the Ranvier nodes.

If one of these electrodes is now activated by a cathodic stimulus, the activity will be propagated, as expected, in a saltatory manner from electrode to electrode in both directions along the glass tube, by means of successive activation releases.

In this model the time course of the propagation process can conveniently be observed by recording both the currents flowing through the "internodal" sections and the potentials at the "nodes." Fig. 16

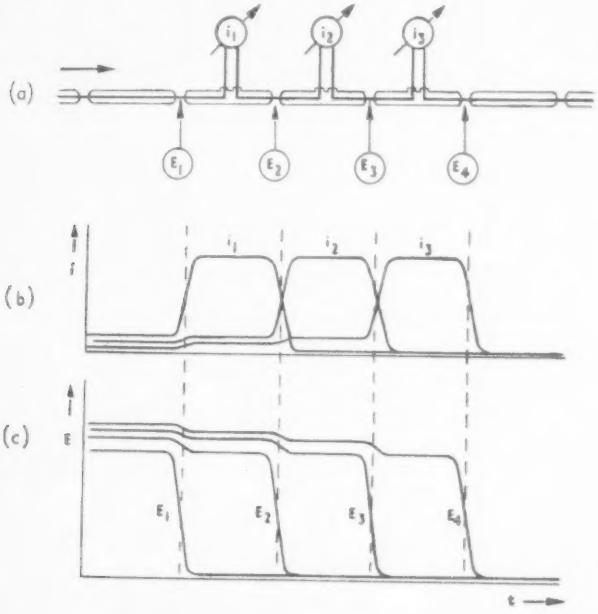


Fig. 16. Current and potential graphs recorded during the propagation of the activation in a saltatory nerve model. (*a*) Location of the reference electrodes; (*b*) change in current intensity with time in the internodal distances; (*c*) change in potential with time at the electrodes.

shows such an oscillogram taken during the passage of a saltatory activation wave. The propagation takes place because the above-mentioned potential difference of about a volt exists between an active and a neighbouring passive node. Consequently between the nodes a current, cathodic at the passive electrode, flows through the axon and back through the outer electrolyte. Thus the electrode will be activated after the passage of a certain latency period which is dependent on the strength of the current. When this occurs the same process commences at the next node. In this way the activation proceeds successively, one internodal distance at a time, whereby the total propagation time is the sum of all single latency periods.

In general the propagation velocity of the saltatory model is greater than that of the continuously conducting Ostwald model. This agrees with the fact that the myelinated nerves conduct excitation 10–100 times faster than the non-myelinated. In the model the velocity will be determined by the size of, and the distance between, the electrodes.

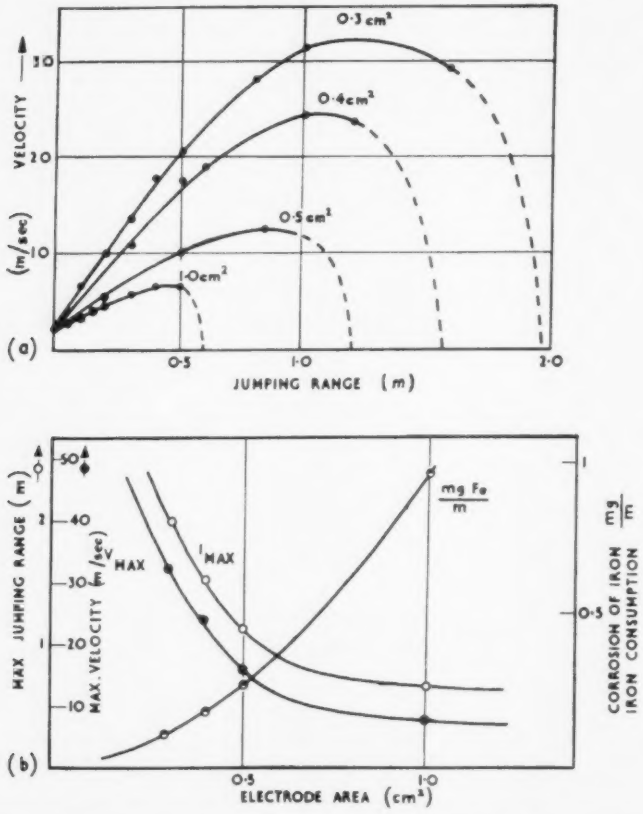


Fig. 17. (*a*) Dependence of propagation velocity in saltatory models on distance and size of the electrodes, plotted from data obtained with models as shown in Fig. 15c. (*b*) Maximum jumping range b_{max}, maximum propagation velocity v_{max}, and iron consumption per propagation over unit distance (1 m).

Fig. 17*a* exhibits the results of velocity measurements on models where both of these determining quantities have been varied. The distance between electrodes is given as the abscissa, the propagation velocity as ordinate, and the electrode surface as parameter of the different curves. All curves commence with a value found for the continuous type of propagation and ascend with different degrees of steepness depending on the size of the electrodes. It is rather striking to observe that the

distance between electrodes cannot arbitrarily be increased but that the velocity, after passing through a broad maximum,[61] decreases again finally becoming zero at l_{max}. This obviously shows that the maximum jumping range is determined by the resistance of the internodal circuit which, on becoming too large, will reduce the current to sub-rheobasic values.[63], [64] Therefore one can speak here of a rheobasic range.

Is noteworthy the large significance which the magnitude of the electrode possesses for the maximum obtainable velocity of propagation. This is larger the smaller the electrode surface area, because the rapidity with which the activation of an electrode proceeds is not dependent directly on the intensity but rather on the density of the current. The latter is produced by the potential difference between two neighbouring electrodes at a given internodal resistance. However the current density is greater the smaller the surface area. Actually the Ranvier nodes are remarkably small as compared with the internodal areas (approximately 5 μ^2 against 10^5 μ^2, cf. Fig. 15*b*). The models also permit statements to be made about the relative energy requirement which is necessary in the case of different "nodal" arrangements for the excitation transfer over a certain distance. Just as the nerve requires a metabolism providing energy (for the propagation of excitation), so the model requires a certain amount of chemical energy for the propagation of activation. After each activation a certain amount of iron must be oxidized (approximately 0·45 mg/cm^2) and an equivalent quantity of nitric acid reduced for the return of iron to the original passive state. Likewise, since in the case of nerve the energy requirement is assumed to be proportional to the excited membrane surface, one can directly relate this to the iron consumption of the model. The maximum jumping range and maximum velocity both of which depend on the size of the electrodes are given in Fig. 17*b* on the basis of the data from Fig. 17*a*. In addition, the resulting loss of iron per activation transmission over unit distance (1 m) is also shown. This diagram clearly shows how, in every respect, the conditions for the propagation of excitation improve with decreasing "nodal" surface area. Thus, the maximim obtainable velocity and similarly the maximum "node" distance which the excitation can overcome will increase; simultaneously, the energy requirement decreases more and more. Since the ultimate purpose of nerve propagation is the fast and safe transmission of information rather than high energy output, the organism proceeds more economically when it permits the excitation to be propagated, not continuously, but rather in a saltatory manner, with high velocities over great distances, and affecting excited membrane areas as small as possible. For that reason, the myelinated nerve, with its long internodal stretches and its extremely small Ranvier nodes, represents, from the functional and economic

point of view, the most favourable solution in the sense of a fast, safe and energy saving signal transmission.

(b) *Generation of high voltage in the electric organs of electric fishes*

With the help of saltatory nerve models, as described above, one can now reproduce any nerve structure and experimentally test its function. As an example we shall briefly discuss a model of the electric organs of electric fishes.[65]

All experimental evidence confirms that also, in the case of the generation of high voltages, the excitation reaction of the protoplasm is the voltage-producing elementary process.

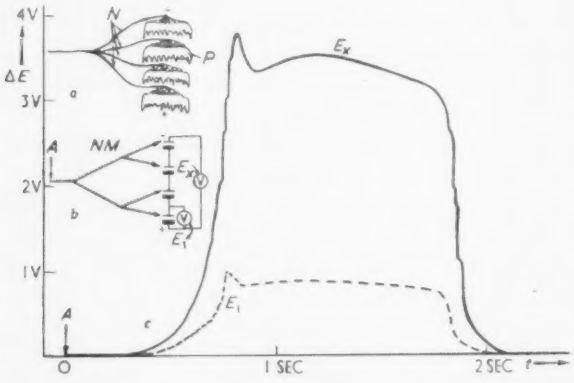


Fig. 18. Model for the voltage summation in the electric organs of electric fishes. (*a*) anatomical structure of the electric organ; (*b*) model design; (*c*) oscillogram showing the summation of four activation waves, yielding the fourfold voltage of one single activation process. *N*: innervation of the electric cells; *P*: electric cells; *NM*: saltatory nerve models.

Essentially the electric organs consist of a great number of regularly arranged electric cells in series, each of which is connected on one side with a neural net[66–73] as is schematically shown in Fig. 18*a*. Each electric cell is able to produce a potential difference of approximately 100 mV for a period of 1–2 msec during the action initiated by the nerve. By means of a series connection and synchronization of single action potentials, the accumulation of the latter obviously makes it possible for the animal organism to attain high voltages of the order of 400–600 V for a short period of time. The nerve system of the electric organ is evidently specialized for the synchronization of single excitations. In the case of the electric catfish (*Malopterurus electricus*) the innervation of the organ, situated in the skin of the animal, originates in only two ganglion cells, one for each side of the body. The myelinated axon of the cell splits up, dichotomously, into as many branches as there are

electric cells (approximately 10^6 in one-half of the body) so that the ganglion cell is placed in direct contact with each electric cell.

The construction of the model closely follows this anatomical design. Here the electric organ is made up of a chain of galvanic elements each of which consists of two passive iron electrodes immersed in concentrated nitric acid. In its resting condition the chain does not generate a potential difference. The one-sided innervation is obtained by saltatory nerve models, all of which originate by regular branching, from one "nerve fibre" (Fig. 18*b*). If the model is stimulated cathodically at the origin A an activation wave will proceed from this point over all branches of the "neural net," to the single electric cells, and produce at each one, for a short time, a potential difference of about 1 V. If the propagation times in the individual branches are equal, so that the activation wave reaches all the electric cells at the same time, these cells will simultaneously generate their action potentials. Thus, between both ends of the galvanic chain corresponding to the electric organ, the sum of all the single potentials will result. In Fig. 18*c* an oscillogram shows this process for the summation of four electric cells. This arrangement can be extended as desired. Owing to the strong divergence of the geometric progression $a_n = 2^n$, one would already in the case of $n = 10$ dichotomous branchings, find a voltage of 1000 V. For further details about these model experiments the original literature must be referred to.[65]

(c) *Electrochemical models for rhythmical nerve activity*

Electrodes that can be activated also provide suitable models for the rhythmical phenomena which play an important part in the neural processes of living organisms.[74]

A considerable number of electrode systems are known which, under constant external conditions, display rhythmical behaviour. Among these is the case of passive iron in nitric acid.[41] L. MEUNIER[75] found a system which consists of a cobalt electrode in contact with a mixture of chromic and hydrochloric acids, and which executes extremely uniform, spontaneous voltage oscillations in the absence of an external current (Fig. 19*a*).

In principle, such rhythmical reacting electrodes represent electrochemical oscillators which can be coupled with one another as in the case of electric oscillatory circuits.[76] This can be achieved galvanically, capacitively, inductively or simply by bringing the electrodes, not in metallic contact, a distance of a few millimetres apart. The coupling then follows as a result of local currents which occur for short times between the active and passive regions of each electrode (during the rhythmical activation and passivation), and which are propagated to a certain extent in the surrounding electrolyte. If the oscillating

electrodes lie sufficiently close to one another, the local currents of one electrode can affect the other.

If the applied coupling is strong, momentary synchronization occurs as a rule (Fig. 19*a*). In the case of looser coupling the synchronization will be incomplete, in a manner characteristic of relaxation oscillations, the sequence of synchronized oscillations being interrupted at regular

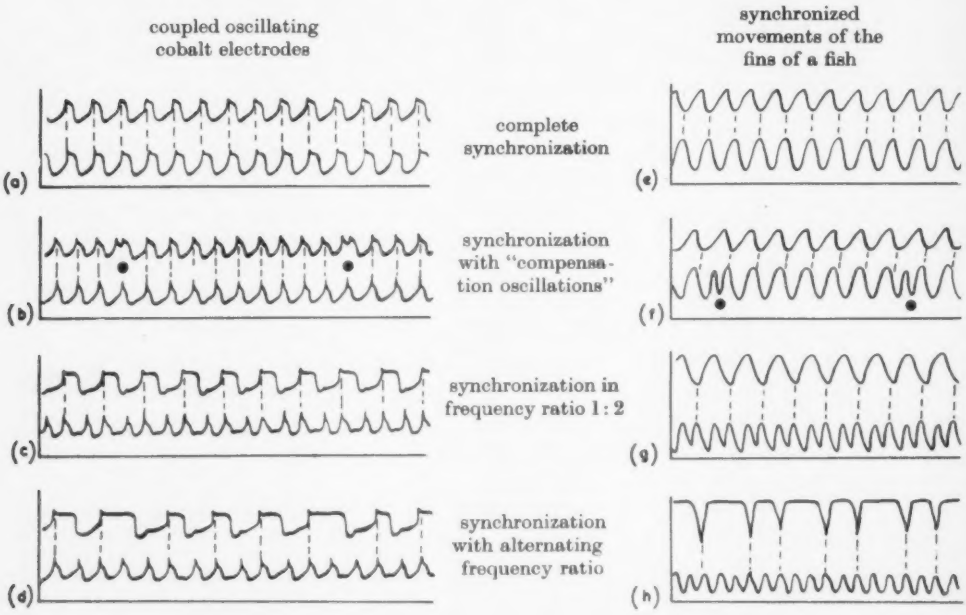


Fig. 19. Comparison between synchronization curves of coupled oscillating cobalt electrodes (*a*, *b*, *c*, *d*) and biological coordinated motion rhythms (*e*, *f*, *g*, *h*). (The oscillograms on the left side are reproduced from E. v. HOLST, *Naturwiss.* 25 (1937) 625, those on the right side from U. F. FRANCK and L. MEUNIER, *Z. Naturforsch.* 8*b* (1953) 396.)

time intervals by "compensation oscillations." These are more frequent the looser the coupling and the greater the difference between the individual frequencies of the coupled electrodes (Fig. 19*b*). If the ratio of the individual frequencies exceeds 2 : 1, the electrodes synchronize in ratio of appropriate integers (Fig. 19*c*), or, on occasion, with a frequency ratio changing periodically (Fig 21*d*),

There is a remarkable agreement between the above described oscillation shapes and those encountered in living organisms. In his study of the fin movement of fishes, E. v. HOLST[77] obtained synchronization oscillograms which, in principle, are completely analogous to those obtained from coupled electrodes, as might be recognized from the comparison in Fig. 19.

The rhythmical movements recorded by v. HOLST are not autonomous oscillations of the fins themselves, but rather of a central nervous origin. The reason seems to be, that, in the spinal cord of the animal, rhythmically reacting ganglion cells send out continuous impulses to the appropriate peripheral motor apparatus of the fins in this manner, determining their stroke frequency. The fact that the frequency of several fins can be completely or partially synchronized shows that interactions must be present between the pulsating ganglion cells. In

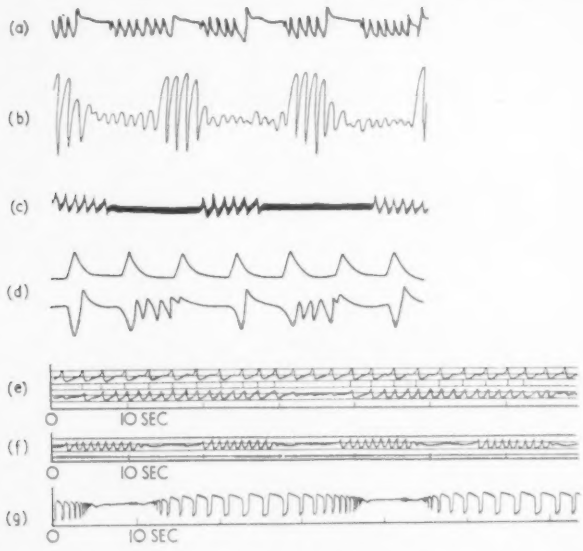


Fig. 20. Rhythmical oscillation groups on living and inanimate relaxation oscillation systems. (*a*) Cheyne-Stokes respiration of a lizard (after Hertel, reproduced from A. BETHE, *Naturwiss.* 33 (1946) 86); (*b*) rhythmical oscillation groups in frog *sartorius tetanus* (after NEURATH, *Pflügers Arch.* 254 (1951) 21); (*c*) periodic action potential of Luciani type measured on the green alga *Nitella* (after OSTERHAUT and HILL, *J. Gen. Physiol.* (1938)); (*d*) rhythmical oscillation groups due to coupling of neon lamp circuits; (*e*) rhythmical oscillation groups of two coupled oscillating cobalt electrodes in chromic acid/hydrochloric acid; (*f*) rhythmical oscillation groups on a single cobalt electrode; (*g*) rhythmical oscillation groups on an anodically polarized iron electrode in sulphuric acid/hydrochloric acid. (*e–g* reproduced from FRANCK and MEUNIER, *Z. Naturforsch.* 8b (1953) 396.)

the most elementary form the stimulating impulses may be considered as influencing and synchronizing each other either directly by local currents—corresponding to the local current coupling between oscillating electrodes—or through synaptic junctions between the ganglion cells.

The rhythmical activity of living cells is often observed to consist of periodically repeated discrete groups of pulses[78] rather than of

uniform oscillations. Well known examples are the Cheyne-Stokes respiration (Fig. 20*a*), the so-called Wenkebach periods accompanying certain heart diseases, the rhythmical tetanus muscle contractions (Fig. 20*b*), the Luciani periods shown by frog heart preparations, etc. Fig. 20 gives a comparison of some characteristic phenomena of this kind with oscillograms obtained from neon lamp circuits and electrode arrangements. This shows that, here also, we have a distinct analogy in behaviour between living cell and inanimate model.

Rhythmical electrobiological processes seem to occur in great variety also in the living brain. As is known, one can derive them as potential oscillations in the form of electroencephalograms taken from outside the head.[79] It is quite probable that these also will be produced by

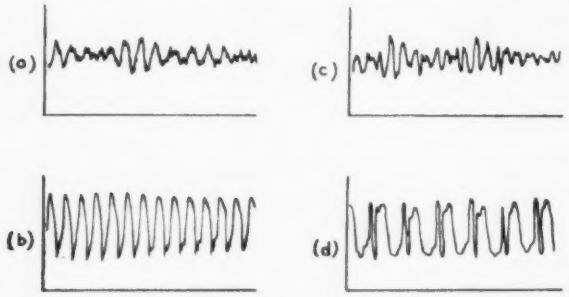


Fig. 21. Comparison of potential oscillograms obtained with an aggregate of ten oscillating cobalt electrodes (*a* and *b*), and electroencephalograms (*c* and *d*).

rhythmically reacting cells. Normally electroencephalograms possess a very complicated appearance which points to the fact that rather complex processes are involved. It is possible that here model experiments with electrochemical oscillators may be of some use.

The simplest concept that we could formulate about the origin of the electroencephalogram would perhaps be to assume that in the brain a great many essentially independent, oscillating, close-lying cells exist. A model corresponding to this could be built from a large number of oscillating electrodes, lying close together, but not in contact with each other.[34] An electrode aggregate of this kind yields a potential diagram of the type reproduced in Fig. 21*a*, and is somewhat similar to a normal electroencephalogram (Fig. 21*c*). Under certain conditions such polyelectrode aggregates may pass over abruptly to synchronization; by changing the composition of the electrolyte the amplitude of the individual oscillators can be raised to such an extent, that the local currents exceed the intensity required for synchronization. When the local currents are not far from having this intensity, total synchronization can also be achieved by synchronizing a few electrodes

through direct galvanic coupling. This seems to intensify the synchronizing effect on the other electrodes, this effect increasing with increasing number of electrodes already synchronized. Accordingly, total synchronization is initiated in autocatalytic manner. The resulting oscillogram shows very regular oscillations of maximum amplitude (Fig. 21*b*).

This synchronization behaviour of the electrode aggregates is of special interest in connection with electroencephalograms observed during certain pathological convulsions.[79], [80] For instance, in the case of epilepsy, closely corresponding oscillograms are recorded, indicating that here also, analogous synchronization processes are involved. This suggestion is supported by the fact that epileptic condition usually commences in the form of an abrupt attack, i.e., autocatalytically. The model can naturally give no information regarding the extent to which the coupling actually follows through the primary rhythmic action currents of the cells concerned or whether synaptic connections play a transmitting rôle: the result is the same in both cases.

(*d*) *Models for interneural interaction and excitation transmission*

If one allows very weak sub-threshold impulses to act rhythmically on an oscillating electrode, these impulses, in spite of their sub-threshold characteristics, may arrive at a manifest stimulation result in the form of synchronization or frequency adaptation through summation. As established by the work of ECHLIN and FESSARD, and others, a similar behaviour is found in the case of rhythmical sub-threshold stimuli on neural and muscular oscillations. It is likely that interactions of this kind play an important part in neural function, for such a summation of rhythmically repeated sub-threshold stimuli renders possible a stimulus utilization and a transmission of information, respectively, from neuron to neuron independent of the All-or-Nothing Principle.

Also in the resting nerve an effect of sub-threshold stimulating action was proved to exist. BLAIR and ERLANGER[81] and, later, MARAZZI and LORENTE DE NÓ,[82] and KATZ and SCHMIDT,[83] have shown experimentally than an excitation impulse passing along one nerve fibre has a noticeable effect on the neighbouring fibres. For example, in the case of the giant axon the threshold is reduced by about 20 per cent (KATZ and SCHMIDT[83]). The interaction is sufficiently strong to synchronize two impulses initiated on neighbouring fibres. A similar effect was discovered by K. YAMAGIWA[84] when investigating the mutual influence of activation waves in parallel Ostwald-Lillie iron wire models. Independently, K. F. BONHOEFFER and G. VOLLHEIM[85] obtained essentially the same results with model aggregates of this kind.

Analogy experiments with electrochemical models might perhaps prove to be useful for a problem which has not yet been settled, i.e.,

the question whether in neuro-neural and neuro-muscular synapses the excitation is transmitted electrically by action currents or chemically by certain excitation substances.[86] As we have seen, the models can only imitate processes of electrochemical nature but they must necessarily fail where special chemical substances take over the transmission of excitation.

It is known that under certain conditions in artificial synapses (ephapses), the excitation can leap over from one fibre to another.[87] This phenomenon was also demonstrated with the Ostwald-Lillie model by K. YAMAGIWA.[88] Accordingly, the effect of the "pre-synaptic" action currents is sufficient to accomplish a "post-synaptic" excitation.

On the other hand, a considerable body of experimental observations, such as the specific sensibility of synaptic function to prostigmine and curare, seem to indicate that, in the synapses, chemical substances liberated pre-synaptically and initiating excitation post-synaptically rather than action currents, are effective. It remains to be seen, therefore, whether model experiments as described here can contribute to the understanding of synaptic transmission.

VOL
6
195

REFERENCES

(1) BONHOEFFER, K. F.; *Naturwiss.* 31 (1943) 270
(2) BONHOEFFER, K. F.; *J. Gen. Physiol.* 32 (1948) 69
(3) BONHOEFFER, K. F.; *Angew. Chem.* 61 (1949) 324
(4) BONHOEFFER, K. F.; *Naturwiss.* 40 (1953) 301
(5) BONHOEFFER, K. F.; *Angew. Chem.* 67 (1955) 1
(6) BERTALANFFY, L. VON; *Nature* 163 (1949c) 384
(7) BURTON, A. C.; *J. gen. Comp. Physiol.* 14 (1939) 327
(8) PRIGOGINE, I. and WIAME, J. M.; *Experienta* 2 (1946) 450
(9) DE GROOT, S. R.; *Thermodynamics of Irreversible Processes*, Amsterdam, 1952
(10) FRANCK, U. F.; *Z. Physik. Chem.*, Neue Folge 3 (1955) 183
(11) PRIGOGINE, I.; *Étude Thermodynamique des Processes Irreversibles*, Liège, 1947
(12) HODGKIN, A. L., HUXLEY, A. F. and KATZ, B.; *Arch. Sci. Physiol.* 3 (1949) 129
(13) HILL, A. V.; *Proc. Roy. Soc. Lond.* B 119 (1936) 305
(14) GERSTNER, H.; *Pflügers Arch.* 252 (1950) 350
(15) ROTHE, H. and KLEEN, W.; *Bücherei der Hochfrequenz* 5 (1941) 9
(16) VALLE, G.; *Physik. Z.* 27 (1926) 473
(17) KAUFMANN, W.; *Ann. Physik.* 2 (1900) 158
(18) STEIMEL, K.; *Jb. drahtl. Telegr. und Teleph.* 36 (1930) 161
(19) PENNING, F. M.; *Physik. Z.* 33 (1932) 816
(20) POL, B. VAN DER and MARK, J. VAN DER; *Arch. neerl. Physiol.* 14 (1929) 418
(21) BETHE, A.; *Naturwiss.* 33 (1946) 86
(22) GREY WALTER, W.; *The Living Brain*, Duckworth, London, 1954
(23) VAN 'T HOFF, J. H.; *Études de Dynamique Chimique*, 1884
(24) SEMENOFF, N.; *Z. Physik* 48 (1928) 571
(25) JOST, W.; *Explosions- und Verbrennungsvorgänge in Gasen*, Berlin, 1939

(26) VERWORN, G.; *Pflügers Arch.* 167 (1917) 289
(27) LUTHER, R.; *Z. Elektrochemie* 12 (1906) 597
(28) FRANCK, U. F.; *Z. Elektrochemie* (in press)
(29) BARTLETT, I. H. and STEPHENSON, L.; *J. Electrochem. Soc.* 99 (1952) 504
(30) WETZLER, G.; *Schweigers Z. Chemie u. Physik* 50 (1827) 88, 129
(31) HEATHCOTE, H. L.; *Z. physik. Chemie* 32 (1902) 368
(32) HEATHCOTE, H. L.; *J. Soc. Chem. Ind.* 26 (1907) 899
(34) FRANCK, U. F.; *Z. Elektrochemie* 57 (1953) 883
(35) LILLIE, R. S.; *J. Gen. Physiol.* 3 (1920) 107; 7 (1925) 473
(36) LILLIE, R. S.; *Biol. Review* 16 (1936) 216
(37) HERMANN, L.; *Handbuch der Physiologie*, Bd. 2/1 (1879)
(38) BONHOEFFER, K. F. and VETTER, K.; *Z. physik. Chemie* 196 (1950) 127
(39) HODGKIN, A. L. and HUXLEY, A. F.; *J. Physiol.* 109 (1945) 176
(40) BONHOEFFER, K. F.; *J. de Chimie Physique* 51 (1954) 521
(41) MURALT, A. VON; *Die Signalübermittlung im Nerven*, Basel, 1946
(42) BEINERT, H. and BONHOEFFER, K. F.; *Z. Elektrochemie* 47 (1941) 441, 536
(43) BONHOEFFER, K. F., BRAUER, E. and LANGHAMMER, G.; *Z. Elektrochemie* 51 (1948) 29
(44) BONHOEFFER, K. F., HAASE, B. and LANGHAMMER, G.; *Z. Elektrochemie* 51 (1948) 60
(45) BONHOEFFER, K. F. and LANGHAMMER, G.; *Z. Elektrochemie* 51 (1948) 67
(46) VETTER, K.; *Z. physik. Chemie* 194 (1950) 199
(47) VETTER, K.; *Z. Elektrochemie* 55 (1951) 675
(48) VETTER, K.; *Z. Elektrochemie* 56 (1952) 106
(49) BONHOEFFER, K. F. and RENNEBERG, W.; *Z. Physik.* 118 (1941) 389
(50) YAMAGIWA, K.; *Jap. Med. J.* 1 (1948) 390
(51) YAMAGIWA, K.; *Jap. Med. J.* 2 (1949) 217
(52) FRANCK, U. F. and SCHURIG, H.; publication prepared, cf. ref. (5)
(53) FRANCK, U. F.; *Z. Elektrochemie* 55 (1951) 154
(54) FRANCK, U. F. and HOPPÉ, K.; to be published, cf. ref. (5)
(55) WOLFROM, I.; Dissertation, Leipzig (1943)
(56) FRANCK, U. F. and LÜDERING, H.; to be published
(57) WEISS, G.; *Arch. Ital. de Biol.* 35 (1901) 413
(58) HOORWEG, P. A.; *Pflügers Arch.* 133 (1910) 161
(59) FRANCK, U. F.; *Z. Naturforsch.* 4a (1949) 378
(60) BLAIR, H. A.; *J. Gen. Physiol.* 15 (1932) 709, 731; 16 (1932) 177; 18 (1935) 755
(61) STÄMPFLI, R.; *Ergebn. d. Physiol.* 47 (1952) 70
(62) LILLIE, R.; *J. Gen. Physiol.* 7 (1925) 473
(63) FRANCK, U. F.; *Z. Elektrochemie* 55 (1951) 535
(64) YAMAGIWA, K.; *Jap. J. Physiol.* 2 (1951) 79
(65) FRANCK, U. F.; *Z. Naturforsch.* 7b (1952) 220
(66) ROSENBERG, H.; *Hdb. d. normalen u. pathol. Physiologie*, Bd. 8/2 (1928) 876
(67) GARTEN, S. and WINTERSTEIN, H.; *Hdb. d. vergl. Physiologie*, Bd. 3/2 (1910) 170
(68) COATES, C. W., COX, R. T. and GRANATH, L. P.; *Zoologica (New York)* 22 (1937) 1
(69) COX, R. T., ROSENBLITH, W. A., CUTLER, J. A., MATHEWS, R. S. and COATES, C. W.; *Zoologica (New York)* 25 (1940) 553
(70) CHAGAS, C. and PERCIRA, L. R.; *J. cell. comp. Physiol.* 28 (1946) 489
(71) FESSARD, A.; *Ann. New York Acad. Sci.* 47 (1946) 501
(72) ALBE-FESSARD, D., CHAGAS, C., CONCEIRO, A. and FESSARD, A.; *J. Neurophysiol.* 14 (1941) 243

(73) KEYNES, R. D. and MARTINS-FERREIRA, H.; *J. Physiol.* 119 (1953) 315

(74) FESSARD, A. and ARVANITAKI, A.; *Proprietés Rhythmiques de la Matière Vivantes*, Paris, 1936

(75) MEUNIER, L.; *Comptes Rendu, Comité Intern. de Thermodynamique et de Cinétique Electrochimiques* 3 (1952)

(76) FRANCK, U. F. and MEUNIER, L.; *Z. Naturforsch.* 8b (1953) 396

(77) HOLST, E. VON; *Naturwiss.* 25 (1937) 625, 641

(78) BETHE, A.; *Allgemeine Physiologie* (1952) Berlin, Göttingen, Heidelberg

(79) HILL, D. and PARR, G.; *Electroencephalography*, London, 1950

(80) KORNMÜLLER, A.; *Klinische Wochenschrift* 31 (1953) 228

(81) BLAIR, E. A. and ERLANGER, J.; *Amer. J. Physiol.* 101 (1932) 559

(83) MARAZZI, A. S. and LORENTE DE NÓ, R.; *J. Neurophysiol.* 7 (1944) 83

(83) KATZ, B. and SCHMIDT, O. H.; *J. Physiol.* 97 (1940) 471; 100 (1949) 369

(84) YAMAGIWA, K.; *Jap. J. Physiol.* 1 (1950) 40, 48

(85) BONHOEFFER, K. F. and VOLLHEIM, G.; *Z. Naturforsch.* 8b (1953) 406

(86) ROSENBLUETH, A.; *The Transmission of Nerve Impulses*, London (1950)

(87) JASPER, H. H. and MONNIER, A. M.; *J. cell. comp. Physiol.* 2 (1938) 259

(88) YAMAGIWA, K.; *Jap. Med. J.* 1 (1948) 452

7

PHYSICAL ASPECTS OF THE SENSE ORGANS

Hl. de Vries

CONTENTS

	Page
INTRODUCTION	208
I. THE SENSES OF TASTE AND SMELL	208
II. THE MECHANICS OF THE ORGANS OF THE LABYRINTH AND THE LATERAL LINE	210
1. Introduction	210
2. The mechanics of the otoliths (theoretical)	213
3. The mechanics of the otoliths (experimental)	214
4. The otoliths as detectors of soundwaves	216
5. The mechanics of the semicircular canals	219
6. Discussion of the equation	222
7. The optimal dimensions of the canals	225
8. The system of three canals	225
9. The mechanics of the organ of hearing	226
10. The middle ear	228
III. THE SENSE UNIT OF THE LABYRINTH ORGANS	229
1. The microphonic effect	229
2. The functional significance and the origin of the microphonics	233
3. The sense unit in the other labyrinth organs	235
IV. BROWNIAN MOTION AND THE LABYRINTH ORGANS	236
1. Introduction	236
2. Brownian motion and hearing	237
3. Comparison of the Brownian motion of the eardrum and the absolute threshold of the ear	239
4. General discussion of the Brownian motion in the sense cells.	240
5. Brownian motion in the inner ear	242
V. PHYSICAL ASPECTS OF VISION	245
1. Introduction	245
2. Effects of diffraction for one single sense cell with a lens	246
3. The optimal construction of a compound eye	248
4. The sense cells as waveguides	250
5. The detection of polarized light	252
6. Lightquanta and vision	256
7. Effects of thermal agitation	258
REFERENCES	263

OL.
6
956

7

PHYSICAL ASPECTS OF THE SENSE ORGANS

*Hl. de Vries**

Introduction

In contrast to biochemistry, biophysics is not yet one homogeneous field; it is still subdivided into nearly as many small parts as there are scientists working in it. The only common feature is that advanced physical methods are being applied to biological problems, and very often by the physicists themselves. This leads to the question why such an increasing number of physicists are becoming interested in biological problems. Perhaps the answer is contained in the fact that the main interest of physicists is now in nuclear physics, biology "being an area encircled by the tank divisions of physics." All biological processes are processes or chemical reactions in which only the outer parts of the shell of electrons are concerned and their description should be completely given by quantum mechanics. This does not imply, of course, that the problems are always easily solved. The conviction, however, that they do not need new theories is a stimulus for their attack and for the attempts to assimilate biology into physics.

Amongst the biological systems the sense organs are especially attractive from a physical point of view by their similarity with instruments. Up to now, the study of the senses of touch, heat and pain has not lead to many details of physical interest. Much more can be told about vision and about the organs of the labyrinth as mechanical detectors, and there is some information about the senses of taste and smell.

I. The Senses of Taste and Smell

Since it is the more sensitive, we shall confine ourselves here to the sense of smell. The sense cells are situated on the mucous membrane of the nose, which, because of its large number of folds (the conchae) has an enormous area, especially for animals with a good developed sense of smell. In man there are only three conchae in each half of the nose and, the olfactory epithelium covers only a small area (see Fig. 1). If we stretch the folds in the membrane, keeping the distance from the walls constant, the lumen of the human nose can be represented by the model

* Though it was my aim to give a general survey of the subject, I have partly followed the lines of our own researches. Moreover, I have included various unpublished investigations (e.g. the work on the insect eye (Section V)). Therefore I wish to acknowledge the co-operation of all members of the biophysical group, and especially its biologist, Dr. J. W. Kuiper.

of Fig. 1. For calculations or experiments on the flow of air this model is very suitable but it should be kept in mind that there are large individual differences. Moreover the lumen of the nose can vary considerably in less than a minute by swelling or shrinking of the conchae by control of their blood supply.

An important question is what part of the odorous substance that is inhaled really comes into contact with epithelium. In order to answer this one has first to find out what fraction of the air really passes along the olfactory epithelium. This already involves a difficult problem since the flow, though laminar, is not governed by the laws of Poiseulle.

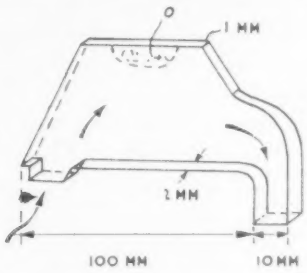


Fig. 1. Schematical representation of the lumen of the human nose. (Partly after PROETZ.[1]) The area O of the olfactory epithelium is about 2·5 cm².

Inertia plays an important part so that a larger fraction of air flows along the epithelium than that which is calculated neglecting the inertia. In "sniffing" this effect is even more pronounced which may explain part of the increase of sensitivity obtained by sniffing. The diffusion of the molecules to the epithelium is also an important factor. One finds that for a rate of flow somewhat below the rate for normal inhalation, nearly all molecules hit the epithelium at least once during the passage along it. The likelihood of reaching the epithelium decreases with increasing rate of flow; this negative effect may be more than compensated, however, by the fact that the loss by absorption on the mucous membrane before arrival at the olfactory slit (according to calculations by M. STUIVER in our laboratory) decreases at the same time.

After colliding with the surface of the olfactory epithelium the molecules have to pass through the thin layer of fluid (essentially water) by which it is covered and finally, if they have to really enter the cells, they have to cross the lipoid membrane. The classical investigations of ZWAARDEMAKER[2] already gave evidence that good solubility in water and fats produces a tendency for a low olfactory threshold. This rule is only a statistical one, however, and it was soon recognized that more individual properties of the molecules were of ultimate importance. Even two optical isomers can have a different

smell. As an example we mention *d* and *l* lemonene, the first smelling like freshly crushed lemon rind, the second like a mouldy lemon.

Though much work has been done on the relation between molecular constitution and smell, this problem is still far from being solved. Moreover, it is a more a problem of biochemistry than of biophysics. A rule, which seems to be more promising than ZWAARDEMAKER's rule, was recently found by DAVIES and TAYLOR.[3] They discovered a close parallelism between the activity of a series of substances in accelerating haemolysis and their efficiency as an odorous substance.

For a check on any theory on olfactory stimulation it is important to know whether a macroscopic response can be started by one single odorous molecule or whether a "macroscopic" amount of material is required so that the rules of "macro-chemistry" can be applied. For human smell the estimates of this minimum number of molecules required varies from about 1 to 10^7. Taking into account the physical phenomena in the nose, mentioned above, STUIVER calculated from his threshold measurements on some mercaptans that the number of molecules per sense cell was of the order of unity. For the nose of a dog it is even easier to prove that one cell is stimulated by one molecule. NEUHAUS[4] found for various fatty acids a threshold corresponding to about one molecule per $3 \cdot 10^5$ sense cells. According to our analysis of the flow of the air inhaled some corrections have to be applied, but even then it is about 1 molecule per 10^4 sense cells, so that the probability that one cell receives more than one molecule is negligible. The same paper also shows that this low threshold makes it possible to understand how a dog can follow traces, even as much as one day old.

The shortness of this section indicates that this field is still expanding and that it deserves the interest of physicists.

II. THE MECHANICS OF THE ORGANS OF THE LABYRINTH AND OF THE LATERAL LINE

§ 1. *Introduction*

Of all the organs of the labyrinth, the ear, at least for human beings, is by far the most important. We shall treat the organs from one general point of view, however. This is possible since they all derive from the same "unit," the sense organs of the lateral line of fishes. These organs consist of a group of sense cells with hairs, the so-called haircells, which are covered with a jelly-like cap, the cupula (Fig. 3). This cupula projects into the surrounding water and consequently water movements will cause a *bending* of the cupula as well as *displacements* of the cupula as a whole relative to its base, the sense cells. It will be shown later that it is the *displacement* which is of functional importance; we shall not therefore mention further the bending of the cupula.

VOL.
6
1956

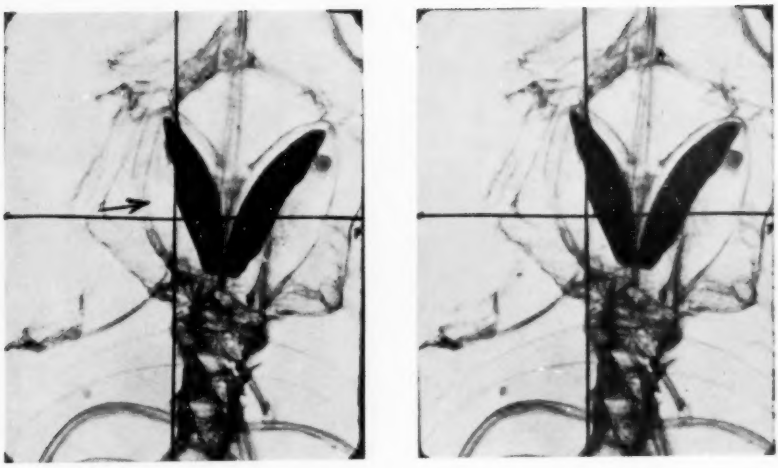


Fig. 4. X-ray photographs of a *ruff*. *Left*: head up. *Right*: head down. Note the displacement of the large sacculus otolith relative to the bony wall of the sacculus (or the crosswires). (From DE VRIES[6].)

VOL
6
195

If we provide this detector of displacements with an otolith, we have constructed a system which can detect the position of the organ relative to the vertical (Figs. 3 and 4).

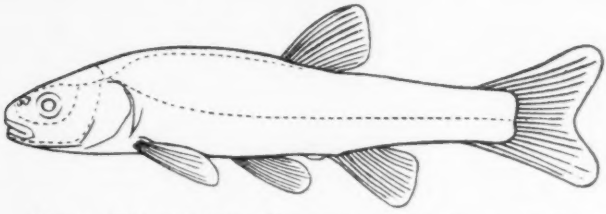


Fig. 2. Fish with lateral line system, after DIJKGRAAF.[5]

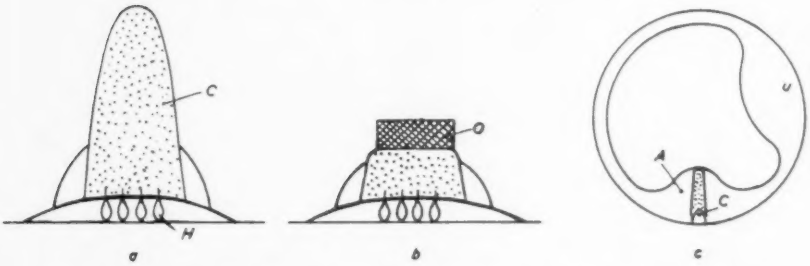


Fig. 3. (*a*) Lateral line organ with cupula *C* and haircells *H*. (*b*) Otolith organ with otolith *O* on the same sense-unit (schematical). In the actual organ the otolith is embedded in the cupula. (*c*) Semicircular canal. The wide part *U* is the utriculus (see Fig. 6). The sense unit (the crista *C*) is in the ampulla *A*, where the cupula closes the lumen like a rotating door.

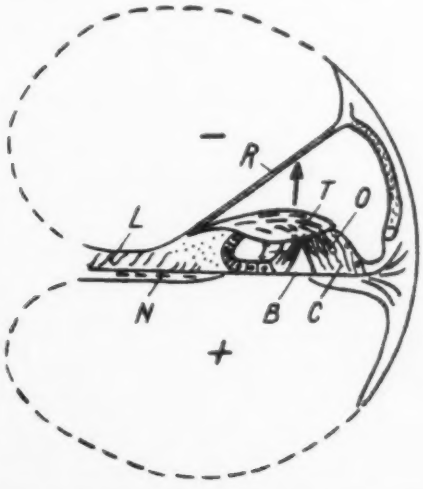


Fig. 5. Cross section of the cochlea. *B*: basilar membrane; *T*: tectorial membrane; *O* and *I*: outer and inner haircells, in the organ of Corti *C*.

A somewhat more complicated mechanical adaptor makes it possible for the same unit to detect rotations (Fig. 3*c*). If this semicircular canal is rotated (the other half of the circle is completed by the wider *utriculus*) the fluid lags behind, and this again, also displaces the cupula in the ampulla relative to its base.

In the cochlea, finally, the cupula reappears as the tectorial membrane (Fig. 5). If sound falls on the eardrum, the vibrations are

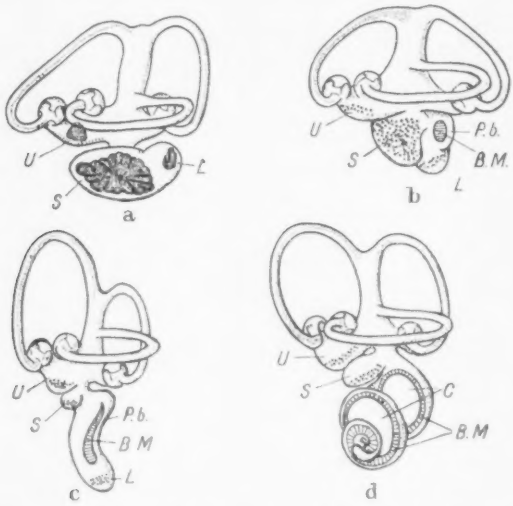


Fig. 6. The complete labyrinth of various vertebrates. (*a*) fish; (*b*) tortoise; (*c*) bird; (*d*) mammal. The upper part (static organs) consisting of semicircular canals and utriculus *u*, are nearly the same throughout the whole system of vertebrates. *s*: sacculus; *l*: lagena; *BM*: basilar membrane of the organ of hearing. (From v. FRISCH.[8])

conducted to the cochlea and produce vibrations (in the direction of the arrow) of the basilar membrane, at least in the middle, since the ends are connected to the walls. By this bulging up and down a shearing motion of the cupula (here the tectorial membrane) results. Fig. 6 gives a general survey of the complete labyrinth of various vertebrates. We can summarize this introduction by the scheme of Fig. 7.

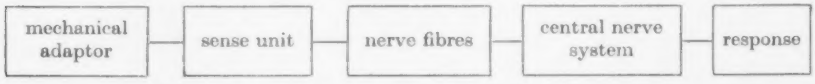


Fig. 7. Schematical diagram of the labyrinth organs.

In the present article we shall not deal with the response which has been recorded from the nerve fibres or from the central nervous system, since this reaction is of interest for the present purpose only where measurements of the thresholds of the organ are being used, in

the discussions. The mechanical adaptor is of obvious interest for the physicist, and because of its extremely high sensitivity the same is true for the sense unit.

§ 2. *The mechanics of the otoliths (theoretical)*

The equation of motion of an otolith is probably simply that of a harmonic oscillator (see also below):

$$m_e\ddot{x} + b\dot{x} + fx = ma(\rho - 1)/\rho \qquad \ldots.(1)$$

The mass m_e is the effective mass, i.e. the actual mass plus the inertia of the water that has to be displaced if the stone moves. We can assume that this extra mass is about the mass of the water corresponding to the volume of the stone, m/ρ, if m is the real mass and ρ the specific weight. The value of ρ depends on whether the calcium carbonate of which the stone consists has the calcite of the aragonite structure ($\rho = 2{\cdot}93$ and $2{\cdot}71$ respectively). For fishes ρ is generally $2{\cdot}93$. The measurement of m, b and f will be described in the next section.

The right-hand side of the equation (1) contains the driving force. Since we are interested in displacements of the otolith *relative to the epithelium* we choose our coordinate system fixed to the head. If the head is accelerated, virtual (inertia) forces equal to ma will act on all masses m. If gravity alone acts, the force is equal to mg. The force on the stone has to be corrected for the "upward pressure" of the water (Archimedes), and this gives the factor $(\rho - 1)/\rho$ in (1). Inserting the values of m_e and ρ, we finally obtain:

$$\ddot{x} + \frac{b}{m_e}\dot{x} + \frac{f}{m_e}x = \frac{\rho - 1}{\rho + 1}a \approx \tfrac{1}{2}a \qquad \ldots.(2)$$

Perhaps the system is not simply a harmonic oscillator since the friction may be partly friction in the layer between the gliding cupula and the epithelium and partly friction in the jelly-like substance of the cupula itself. In the same way the elasticity may be partly localized in the cupula itself, and be partly due to the structures that connect the cupula to the epithelium. We shall neglect this complication, however, since it need not have important influence on the final results we are interested in.

If an acceleration a is applied, or if the head is tilted, the stone will move to a new equilibrium position. It follows from (1) that the final displacement is

$$x = (\rho - 1)/\rho \,.\, ma/f \qquad \ldots.(3)$$

From this formula f can be calculated if the other constants have been measured.

Before the otolith comes to rest in its new position, however, it can perform oscillations in which it overshoots the new position; this

depends on the value of b. In all modern instruments, such as voltmeters, etc., the damping b is so chosen that these vibrations are suppressed and this is obviously also desirable for the otolith system. b should not be larger than the *critical* value since this would imply a slower motion towards the new position. The "critical" value of b is equal to $2\sqrt{m_e f}$. The *time of indication* τ for this value of b is about $1/n_0$, n_0 being the resonance-frequency. Summarizing: $\tau = 2\pi\sqrt{m_e/f}$.

Response to vibrations. It is obvious that the system will also respond to vibrations of the head, since the stone tends to lag behind, giving a displacement relative to the epithelium. According to PUMPHREY[7] this is its prime function. If the amplitude of the head is u_0, the acceleration is $u_0(2\pi n)^2$. The amplitude u of the response can be calculated from (2); one finds

$$(u/u_0)^2 = 1/4[(1 - (n_0/n)^2)^2 + (1/2\pi n)^2 b^2/m_e^2] \qquad \ldots.(4)$$

For the resonance frequency the quotient u/u_0 only depends on b, so that conversely, b can be calculated from the response (see next section).

For high frequencies, the quotient u/u_0 becomes $\frac{1}{2}$. In the resonance point it depends on b, the value for critical damping being $\frac{1}{4}$. For low frequencies it finally becomes $\frac{1}{2}(n/n_0)^2$. For very low frequencies this amplitude u becomes approximately zero; this means that the stone completely follows the head, as would be anticipated.

§ 3. *The mechanics of the otoliths (experimental)*

We have seen that the otoliths can record the position relative to the vertical (static function) as well as vibrations. The actual effect will depend on the sensory epithelium, which may respond selectively to constant displacements (or low frequencies) or to higher frequencies. There is good evidence that, at least in the higher vertebrates, the utriculus has a static function, whereas the sacculus at least has an acoustic function (see VON FRISCH[8]). The relative positions of the three otoliths are demonstrated in Fig. 6.

Up to now the only quantitative information about f and b (see § 2) which is available was derived from X-ray photographs of surviving heads of fishes (see Fig. 4 and ref. [6]).

The stiffness f could be determined from the displacement by gravitational forces, or by centrifuging if larger forces are required. Fig. 8 gives the relation between displacement and force for the sacculus of the ruff (*acerina cernua*). It is interesting to note that this otolith is "blocked" if the force exceeds the gravitational force. For the smaller otoliths the displacements were smaller and no indication of this "blocking" was found. The differences in f between the various otoliths were not appreciable and we give here only the average value of 1000 dyne/cm.

The friction b has been determined from the response to vibrations (see § 2) but only in the case of the sacculus of the *ruff*. We found $b = 15$ c.g.s. units.

The weights of the three otoliths in the sacculus, utriculus and lagena are very different. The differences between small and large

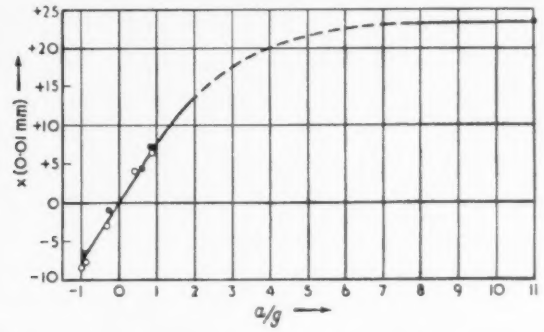


Fig. 8. Displacement of the sacculus-otolith of the *ruff* for various accelerations a. The parameter g = gravitational acceleration (see text).

fishes are very small compared with the difference in weight of the whole fish (see Table I). This is a general rule that holds for all parts

TABLE I

Weights (in mg) of the Otoliths of some Fishes together with the Weight of the Fish (grams) (from DE VRIES[6])

	Fish	*Sacculus*	*Utriculus*	*Lagena*
Pike . .	35	4·4	0·5	0·5
Pike . .	3000	51	6	6·5
Ruff . .	6	7	0·16	—
Ruff . .	27	35	0·65	0·12

of the labyrinth: the dimensions of the labyrinth increase far less than do the dimensions of the animal. This is evidently because the labyrinth is not the rudder of the ship but only its compass. Another interesting fact in this connection is that at least in the case of human beings the labyrinth does not grow after birth although the petrous bone, in which it is embedded, grows considerably. The same is probably true for various other mammals.

The time of indication, defined in the foregoing section, can now be calculated for the sacculus of the *ruff*, the only case where we have measured the damping. It turns out that the damping is nearly

critical; the time of indication being 0·02 sec. Assuming that the other (smaller) otoliths are also critically damped, their time of indication is about 0·005 sec. The time taken by an animal to react to a disturbance of equilibrium will be longer than this time of indication of the otolith, but it is evident that a time of about 0·01 sec. is sufficient for a timely reaction in the case of disturbance of the equilibrium. It should be mentioned here that for human beings the otolith (only the utriculus otolith is left in mammals) is not as important for the conservation of equilibrium as it is for lower animals. The only important effect, which is reported for deaf-mutes (who often also have an impaired static function of the labyrinth), is that they are not subject to motion sickness.

So far we have considered the otoliths as "static" instruments which indicate the position relative to the vertical. It is evident that accelerations of the animal will also displace the otoliths, which tend to lag behind. This force of inertia acting together with gravity and generally the static organ (+ central nervous system) interprets the resulting force (constructed according to the parallelogram of forces) as the direction of the vertical. For example, a man lying on a carriage, interprets a horizontal acceleration in the direction of his feet as a tilting which brings his head lower than his feet. Again, in an aeroplane making a turn the resultant force is perpendicular to the bottom of the plane so that for objects within the plane itself no contradiction exists between information of touch, vision and the labyrinth. Seeing objects outside the plane, e.g. trees and houses, however, produces an alarming sensation of something being wrong.

It is probable that closer investigation will show that it is not exactly the new direction of the resultant force which is indicated by the static organ. Assume that a person is lying in an elevator, which is accelerated. It is certain that the otoliths will be displaced by this *extra force*. One can hardly imagine that this displacement will not give rise to a sensation of being tilted, whatever the position of the head.

§ 4. *The otoliths as detectors of soundwaves*

Especially in fish the otoliths have not only a static function but they may also be used for the detection of sound. A soundwave passage through water will rock a fish which has about the same density to and fro. The otoliths, however, will lag behind in the oscillation and the resulting displacements relative to the sense cells (can) give rise to excitation of the nerve-fibres originating from it. The amplitude of the stone relative to the epithelium was calculated in § 2.

Apart from this system the fish has another possibility for detection of soundwaves, by making use of airbubbles in the skull or by use of the swimming bladder. An airbubble is a region of high compressibility

in the system. Consequently water will flow to it from all sides as to a leak, if a pressure wave passes the bubble. It is more or less a matter of choice how to make use of the large local amplitude at the surface of the airbubble. Some fishes (a.o. the Mormyrides) have an airbubble close to the sacculus which is still connected by a rudimentary tube to

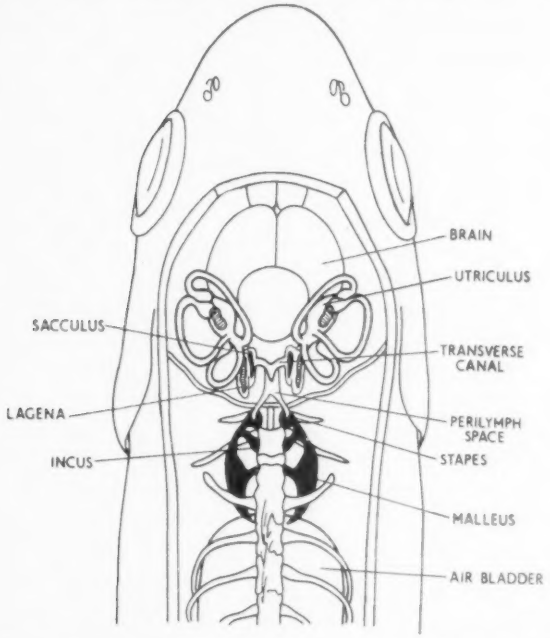


Fig. 9. The coupling between the air bladder and the labyrinth by means of Weber's bones (black). (From v. FRISCH.[8])

the swimming bladder of which it is an offshoot—SCHNEIDER[9] has shown that the hearing is impaired if the air is replaced by water. Another group of fishes, the ostario phsysina, (e.g. the minnow) which have a good hearing, possess a chain of small bones, the bones of WEBER (Fig. 9) which lead from a bony cap on the swimming bladder to the labyrinth. Obviously these are analogous to the ossicles in the middle ear of mammals. It is, however, still possible that they only transmit the hydrostatic changes of pressure with depth. Indeed, DIJKGRAAF[10] has shown that extirpation of the bones impairs the vertical "stabilization" of the fish. On the other hand, fishes which possess WEBER bones do have a low threshold of hearing (VON FRISCH[8]) but, of course, this may be an indirect consequence.

We conclude this section with a quantitative comparison of hearing by otoliths and by airbubbles. The quantitative discussion is only possible for a well-defined model. We assume that the airbubble is

spherical and we calculate the velocity amplitude u at its surface. One finds:

$$u/p = (i\omega r + c)/\{3P - i\omega cr(3P/\omega^2 r^2 - \rho)\} \quad \ldots.(5)$$

where p = pressure of the soundwave, ω = frequency times 2π, r = radius of bubble, c = velocity of sound in water, P = pressure of the air. If u_0 stands for the amplitude of the soundwave, with $u_0 = p/\rho c$, the gain is u/u_0. For this we find from (5)

$$u/u_0 = \{1 + iKE/P\}/\{3 - i(3 - K^2E/P)/K\} \quad \ldots.(6)$$

where E stands for the modulus of elasticity of the water, whereas K stands for $2\pi r/\lambda$, λ being the wavelength. This means that we can neglect K relative to 1 for the frequencies which interest us. The gain is maximum when the factor $3 - K^2E/P$ is zero. Then the frequency n turns out to be $(2\pi r)^{-1}(3P)^{\frac{1}{2}}$. For a sphere with a radius of 1 cm this "resonance frequency" is 280 sec.$^{-1}$: the gain is then 7000. The resonance peak is sharp, the half value width being about 1 per cent of the frequency. Undoubtedly there will be more damping if the bubble is surrounded by the tissues of a fish, and the gain will then be correspondingly less. For higher and lower frequencies one finds from (6) $u/u_0 = \lambda/2\pi r$ and $u/u_0 = 2\pi rE/3P\lambda$ respectively. It can be seen that the gain is negligible at 20,000 per sec. and on the lower frequency side at 4 per sec. (the value of r was taken as 1 cm). Beyond these limits the formulae should not be used.

We see that there is a considerable gain over the whole range of audible frequencies, at least when we consider the amplitude at the surface of the bubble relative to the amplitude far from the bubble. The stimulus for the sense organ, however, is the displacement of a cupula relative to the epithelium. This involves a suitable arrangement of the organ, so that the epithelium remains at rest while the cupula moves with the airbubble. This can be achieved in various ways.

For the otoliths the variation of the response with frequency has already been discussed in Section 2. For all frequencies the response is much smaller than for the airbubble. Though LÖWENSTEIN[11] has demonstrated that nerve discharges from the otolith system can be started by vibrations, it is known that the threshold for fishes with WEBER bones is much lower. Obviously these fishes make use of the second mechanism that responds to the pressure-variations which accompany the soundwave. Moreover, AUTRUM[12] has shown that Amiurus (a fish with Weber bones) "hears" the pressure and not the displacement. He found that in a container with standing soundwaves the threshold was lower in the nodes than between them. In the nodes the pressure variations are maximum; between the nodes the displacement has its maximum.

VOL
6
195

§ 5. *The mechanics of the semicircular canals.* The fundamental property of the system is that the deviation of the cupula from the equilibrium position is proportional to the angular velocity of the head and is independent of the rate at which the velocity is attained—at least under "normal" conditions.

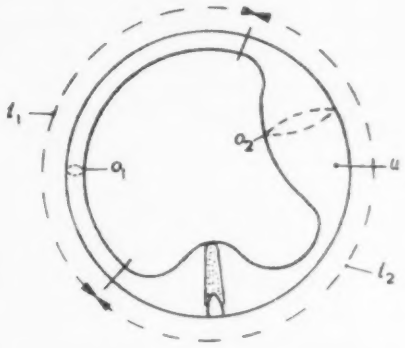


Fig. 10. Semicircular canal. (See text and Fig. 3.)

Though this statement is easily proved in this qualitative form we shall at once derive the quantitative relations.

1. The effect of an angular acceleration α rad/sec^2 is equivalent to the effect of a virtual inertial pressure p

$$p = 2\rho A\alpha \text{ dynes/cm}^2 \qquad \ldots.(7)$$

which drives the fluid through the canal, so that it lags behind the rotation of the head. (ρ = specific weight of fluid, A = area of canal.)

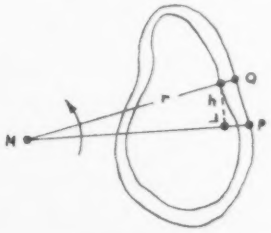


Fig. 11. Semicircular canal (schematical) rotating around M (see text).

Though this is proved most easily for a circular canal which rotates around its centre, this might, however, lead to the common misunderstanding that it is essential to rotate around the centre of the canal.

Considering the part PQ of a canal (Fig. 11) which is rotated around M in the direction of the arrow, an angular acceleration α corresponds to a linear acceleration αr of PQ, perpendicular to the radius. From the point of view of the rotating system, the fluid tries to lag behind and in

this coordinate system the fluid exerts at P a hydrostatic pressure independent of the cross section of the tube, of $h\rho\alpha r$ dynes/cm^2. The product hr is twice the area of the triangle MPQ so that the algebraic sum of all pressure differences will be as given in (1).

2. The (virtual) pressure p will at first accelerate the fluid but the fluid quickly attains a steady velocity relative to the walls, for which the pressure p is completely balanced by the friction. Denoting by p_f the pressure balanced by friction, we have according to POISEULLE

$$p_f = V 8\pi\eta l_1/O^2 \qquad \ldots.(8)$$

The cross section of the wide part of the canal, the utriculus, is at least ten times the cross section O of the narrow part. According to (8) the wide part of the canal hardly contributes to the friction so that we need only take into account the narrow part with length l_1. Instead of the quotient V/O we shall use the *average* velocity v of the fluid in the canal. (The fluid does not all move at the same speed in the canal; it is stationary at the walls and the velocity is a maximum at the centre of the tube—at least there is no indication that its behaviour is different from normal.) Instead of (8) we get

$$p_f = 8\pi\eta l_1 v/O \qquad \ldots.(9)$$

3. During the initial phase of the acceleration a pressure p_i is required to give the fluid an acceleration a, relative to the tube. One finds:

$$p_i = \rho a l \qquad \ldots.(10)$$

In the wider part of the tube the velocity and acceleration of the fluid are proportionally less, consequently only the narrow part of the tube of length l_1 contributes appreciably to p_i. Because of the non-uniform distribution of the velocity over the radius (10) is not quite right; a better approximation gives

$$p_i = \frac{4}{3}\rho a l \qquad \ldots.(11)$$

4. If the fluid flows for a longer time, it will produce an appreciable displacement of the cupula which fits closely in the ampulla, the wide part of the canal. This produces a pressure p_e proportional to the displacement u,

$$p_e = fu \qquad \ldots.(12)$$

Summarizing: $p = p_i + p_f + p_e$ or

$$p = \frac{4}{3}\rho l_1 a + 8\pi\eta l_1 v/O + fu = (l_1 + l_2)^2\alpha\rho/2\pi \qquad \ldots.(13)$$

VOL
6
195

On the right-hand side of the equation we have assumed that the canal is circular, so that we can express A from equation (7) in terms of $(l_1 + l_2)$.

The solution of this equation is straightforward if the values of the constants are known. Unfortunately not all the constants are known with the desired accuracy. The anatomical data can be obtained from various sources but we have taken those of OETLINGER[13] since these measurements were carefully made for use in calculations of the mechanics (see Table II. The data for the pigeon have been taken from WULF[14]). Obviously the dimensions do not vary very much as was emphasized on page 215.

According to TEN DOESSCHATE[15] the viscosity (η) of the endolymphe is not appreciably different from the viscosity of water (η water 0·01 at 20°C and 0·007 at 37°C).

The most serious lack of data is for the stiffness f of the cupula. Some information can be obtained from a photograph published by STEINHAUSEN.[16] The cupula on the photograph was displaced by the "caloric effect," i.e. by a different temperature of the two halves of the canal. Making reasonable assumptions about the circumstances one calculates for the stiffness f from (12) the value of 70 dyne/cm² per centimetre displacement in the canal. For the cupula itself the stiffness

TABLE II

Anatomical Data for the Semicircular Canals. O_c and O_a are the Internal Cross-sections of the Canal and of the Ampulla (in mm²); l_1 and l_2, Length of the Narrow and the Wide Part of the Canal in cm (Fig. 10)

	O_c	O_a	l_1	l_2
Pike (1 kg) .	0·125	0·9	1·2	1·0
Pigeon . .	0·04	0·38	1·4	—
Man . .	0·05	0·8	1·5	0·8

is about 4 dynes/cm; this is only a fraction of the stiffness with which the *otoliths* are coupled to the equilibrium position (about 1000 dynes/cm: see § 3).

Inserting the numerical values in (13) one finds:

$$2\ddot{u} + 520\dot{u} + 70u = 0{\cdot}84\alpha \qquad \ldots.(14)$$

The way in which the formula was derived implies that it describes the motion of the endolymphe (and of the cupula) in a coordinate system that rotates with the labyrinth. Needless to say this is exactly what we want to know.

§ 6. *Discussion of the equation*

Because of the dominating effect of the friction we need not write down the complete solution but we can confine ourselves to some special cases.

If one turns ones head the whole process of acceleration and deceleration lasts a time t which is at most one second. The displacement u in (14) which is of the order of $\dot{u}t$, will consequently be smaller than $\dot{u}$, and we can see that in (14) the term $70u$ can be neglected compared with $520\dot{u}$. The solution of the reduced equation is

$$\dot{u} = \frac{0{\cdot}84\alpha}{520}(1 - e^{-260t}) \qquad \ldots.(15)$$

The equation shows that for t greater than 1/260 sec. the velocity has practically reached the constant value which depends only on the coefficients of the velocity (the friction) and on the coefficient of α.

If the acceleration lasts for a time T the final angular velocity is αT. The final displacement of the endolymphe is $\dot{u}T$. Since, according to (15) $\dot{u}$ and α are proportional, it follows, as has already been stated, that the *displacement* of the endolymphe (and of the cupula) is proportional to the *angular velocity* of the head, except during the first few milliseconds (somewhat more than 1/260 sec.) which will be called the *time of indication* τ. τ is analogous to the time of indication defined for the otoliths. Here again τ is very small. Neglecting time delays of this order of magnitude, the cupula acts as the pointer of an instrument that indicates the angular velocity. The displacement u, according to (15), is equal to 0·84/520 times the angular velocity (in rad/sec.). So we find for 1 rad/sec. (= 57°/sec.) a displacement of 16 microns. According to Table II, the ampulla is 16 times wider than the canal, and here the displacement will be about 1 μ. This is of the same order of magnitude as the adequate displacements in the lateral line (Fig. 17). A wider canal will give a larger displacement.

Cupulometry. The conclusion arrived at above, is only valid if the rotation lasts less than, say, a second but this restriction is not a serious one under the normal circumstances for which nature has built the organ. Physiologists and constructors of Tivoli-gardens, however, have invented rotating machines in which rotation can proceed for longer periods. Under these conditions the elasticity of the cupula can no longer be neglected.

We shall confine ourselves to the simple case in which a certain angular velocity ω is reached within about one second. The velocity is indicated by the position of the cupula in the way we have calculated. If now this constant rotation proceeds (14) reduces to

$$520\dot{u} = -70u \qquad \ldots.(16)$$

This means that the elastic forces move the cupula back to the equilibrium position, but only very slowly because of the viscosity of the endolymphe that has to be displaced through the canal. The solution of (16) is

$$u = u_0\, e^{-70t/520} \qquad \ldots.(17)$$

After about 7 sec. the deviation is only $1/e$ of the initial deviation, etc. After about one minute the cupula will be at rest. If now we stop, the cupula (+ endolymphe) will run ahead in the direction of the rotation and finally attain the same deviation u_0 but in the direction opposite to its first deviation, but the labyrinth does not have the corresponding angular velocity ω. Obviously nature has not anticipated these tricks.

It has been the merit of the Utrecht school of Otologists (physicist J. J. GROEN) that they have recognized that this "abnormal" motion of the cupula is still represented by the sensations which normally correspond to these same deviations. Qualitatively this means that a subject who has been rotating with constant velocity for about one minute thinks he is at rest. If he is stopped, he thinks he is now turning in the opposite direction. In the course of a minute the cupula creeps back again and during this time the experienced sensation of rotation decreases. The innovation was to use trained persons to make a quantitative check of (17) possible, i.e. by having them report the moments when they think they have made one turn, etc. The angular velocities derived from these data fall off exponentially. The coefficient of t in the exponential term is indeed about 1/7, but this agreement is fortuitous since we have little knowledge of the anatomical constants of our own labyrinth.

For untrained subjects the procedure is that they are given various initial velocities and they have to report the moment that they think they come to rest. Assuming that this corresponds to a certain minimum perceptible deviation of the cupula, it is easily shown that there is a similar logarithmic relation between the length of after sensation and the initial velocity. Fig. 12 is an example of such a cupulogram.

The tacit assumption of cupulometry is that the sensation is like a second indicator which moves in exactly the same way as the cupula. Whether this is so depends on the characteristics of the nervous system, however, and there is no reason to assume that nature has made the time-constant of this element in the chain long enough to stand the abnormal test applied here. Obviously the deviations are small, but GRAYBIEL[17] has shown that this second indicator also creeps back after some time. He applied a constant acceleration. After about one minute the cupula reaches a constant position, but in order to keep the sensation constant the acceleration had to increase.

It has been shown that it is possible to find the coefficient of t in (17) from the cupulogram; this coefficient contains the coefficients

of u and $\dot{u}$ in (13). Unfortunately it is very difficult to determine the time of indication (defined above), from the quotient of the coefficients of $\ddot{u}$ and $\dot{u}$. GROEN[18] has recognized that information can be obtained from the resonance frequency of the system. Because of the high damping, the resonance frequency is not in this case the frequency for which the system has a large response, but the frequency for which the phase shift between the cupular displacement and the velocity of rotation is zero. Instead of the cupular displacement, the

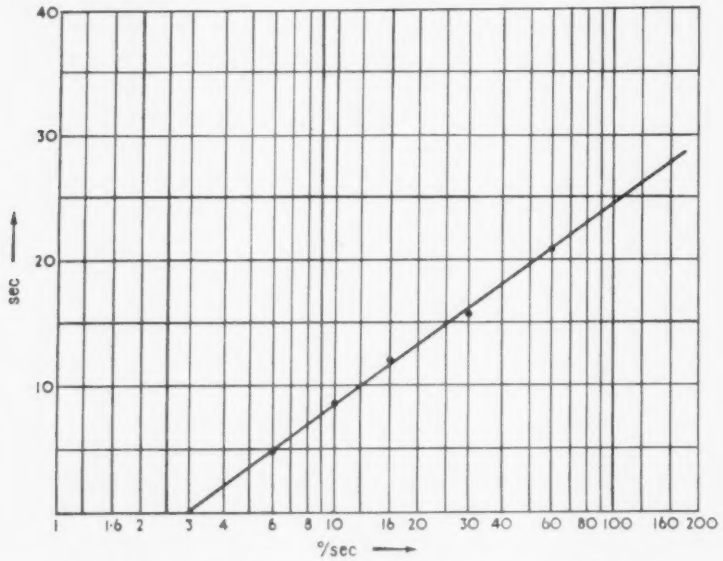


Fig. 12. Cupulogram of a normal observer (see text).

sensation of the observer in the "torsion swing" was used so that all phase shifts between the cupula and the sensation (if they exist) will be included in the experiment. In his experiments GROEN found a resonance frequency of 1/6 per second. The time of indication calculated from this figure is very long (0·1 sec.). In our laboratory J. EMCK has repeated these experiments with a special apparatus using the utmost care. Different amplitudes were used at each of the various frequencies since the phaseshift between the "central" pointer and the cupula might be smaller for larger amplitudes. It was not possible to draw any conclusion with certainty, probably because the times of reaction of the observers were too variable compared with the time differences involved (about 0·1 sec.). HALLPIKE and HOOD[19] also claimed that they confirmed GROEN's results, by a different technique, but after private discussion we agreed that this experiment did not give information on this subject.

§ 7. *The optimal dimensions of the canals*

It follows from the preceding discussion that three parameters are of importance: the time of indication τ, the time of recovery T (in which the cupula creeps back to $1/e$ of its initial deviation, e.g. if it has been treated in an abnormal way by rotations of a long duration), and finally the sensitivity G. G is defined as D/ω, where D is the deviation of the cupula, and ω the angular velocity. In the preceding section we calculated the numerical values of the parameters. In terms of the anatomical data we have:

$$\tau = O_c/6\pi\eta; \quad T = 8\pi l_1\eta/fO_c; \quad G = (l_1 + l_2)^2 O_c^2/16\pi^2\eta l_1 O_a \quad \ldots.(18)$$

It would be of interest to investigate whether the dimensions of the canals are themselves adapted to the special circumstances to which an animal is adapted and also to find out whether the parameters τ, T and G vary if the labyrinth grows—as it does during the life of a fish. No data are available, but considering an animal which makes sudden head movements, τ should be small, and this could be achieved by a narrow canal (see (12)). This would reduce the sensitivity, however, and this could be compensated by a canal of greater length $(l_1 + l_2)$.

One could also suggest making the value of T larger for the human labyrinth which would make it better adapted for longer lasting rotations. This change would also have the effect, however, that the cupula takes longer to come back to its right position when it has been exposed to abnormal conditions. The present value of T seems to be a reasonable compromise for normal conditions.

§ 8. *The system of three canals*

Up to now we have only considered one canal, and an axis of rotation perpendicular to it. If the normal to the plane of the canal and the axis of rotation are at an angle ϕ to each other, the displacements of the cupula, corresponding to the same angular velocity, are multiplied by $\cos\phi$.

It is obvious that one canal can only detect the component of the rotation in its own plane. With very rare exceptions, the vertebrates have in each labyrinth (at each side of the head) three canals (see Fig. 6) in three planes which are nearly perpendicular to each other. From the excitations in the three canals it is possible, in a simple way, to reconstruct the axis of rotations, as well as the angular velocity. This reconstruction is obviously performed by the central nervous system automatically, and we are not aware of the excitations of the individual canals. Nothing is known, however, of the circuits by which the information of the three canals is integrated. We find a similar situation in colour-vision. According to HELMHOLTZ's theory we have three kinds of cones with different spectral distributions of their sensitivity, but we

are not aware of their individual stimulation. Corresponding to only one sensation of the magnitude of the angular velocity we have a sensation of brightness, whereas the direction corresponds to colour. The direction can be given by two parameters, say azimuth and altitude. In the domain of colour this corresponds to the two "directions" of the "four-colour" theory (HERING) red-green and yellow-blue.

Fig. 6 demonstrates that part of the vestibulum is shared by the three canals. The question arises whether this does not give rise to an interference between the canals. Indeed, when the fluid in one canal is propelled by an acceleration, part of it may flow through into another canal. The other canals are, however, a shunt with a high resistance on the path they have in common, so that the leakage is small. Moreover, this effect could easily be taken into account by the computing mechanism of the central nervous system, since it is a linear effect.

§ 9. *The mechanics of the organ of hearing*

The ear of a mammal is a very complicated instrument with a large number of ingenious details of adaptation to the circumstances or requirements. It will not be possible to give a complete survey of the present knowledge. We might refer to VON BÉKÉSY and ROSENBLITH[20] for a recent review.

Because of the difference in acoustic impedance of air and water hearing in air offers special problems which had to be solved by terrestrial animals. For the aquatic animals see § 4. The impedance of a medium for soundwaves is $Z = \rho c$; ρ = specific weight, c = velocity of sound in the medium. At the boundaries between two media a wave is partially reflected, the fraction R of the energy which is reflected, being given by

$$R = (n - 1)^2/(n + 1)^2 \qquad \ldots.(19)$$

where n is the ratio of ρc in the two media. For the boundary air-water n is about 5000 so that less than 1/1000 part of the energy is transmitted. This calculation demonstrates the difficulty in general. In the real ear we are not dealing with two infinite media for which (19) holds so that we shall now have to go into more detail.

Fig. 13 shows the ear, apart from the eardrum and the ossicles of the middle ear. The ear is shown in an uncoiled state as it actually exists in birds. A soundwave will exert equal pressures on the two windows so that only a small compression of the fluids will result. If the sound comes more from either above or below the phase difference between the pressure at the two windows will have a much larger effect since fluid will be forced from one window to the other through the helicotrema (the elasticity of the cochlear partition will be considered later). An even more effective method is to shield one window from the

incident sound: (this occurs in some patients who lack the middle ear). This is the situation which will be assumed to be realized now.

If now sound falls on the oval window which is loaded with a column of water of height $2l$, for a frequency f the impedance of the column (the quotient of driving force and velocity amplitude) is $2l \cdot 2\pi f$ per cm^2. The impedance of air, per cm^2 is $\rho c = 40$ c.g.s. units. Consequently, the window and the air are matched only at very low frequencies, since for the human cochlea l is 3 cm. At higher frequencies a better match can be achieved if an elastic element (i.e. the cochlear partition) is added which compensates for the inertial effect of the mass in a

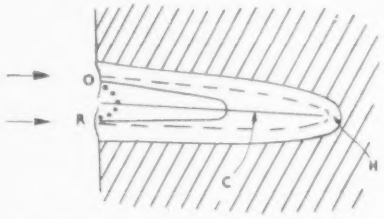


Fig. 13. Schematical drawing of cochlea. C = cochlear partition, with organ of CORTI, etc., see also Fig. 5; H = helicotrema; O and R oval and round window respectively.

region around the resonance frequency—but in the resonance region only. As an illustration we can assume that the fluid then follows the full line in Fig. 13.

The present construction of the ear has the attractive possibility that it can reduce its length for the higher frequencies, the wave following the shorter path indicated by the dotted line. Moreover the stiffness of the cochlear partition in the actual cochlea increases from helicotrema to windows. Unfortunately, the mode of flow of the fluid, suggested above, does not satisfy the equation of motion. Flow will not be restricted to the given lines, but the bubble that is produced on the cochlear partition will itself be propagated as a pulse-wave in an elastic tube. It was a long time before it was realized that the cochlea should be treated as an elastic tube in which waves are generated by the motion of the stapes in the oval window. No essential complication is introduced by the fact that the cochlea is a paired elastic tube, the two halves having the elastic section of the wall in common. A real drawback has been that the constants which had to be inserted in the formulae were not known. They were determined in the well-known work of VON BÉKÉSY.[21–23] Since the constants of the tube, the stiffness and cross section, decrease from windows to helicotrema, and since the friction is considerable, the differential equation and its solution are not simple. Of the different attempts at more or less

complete solutions I mention here the work of ZWISLOCKY.[24] Some of his results are:

1. The input-impedance (at oval window) is nearly independent of the frequency, and real, the numerical value being 9000 dyne-sec./cm^3.

2. The damping is so high that the waves do not reach the helicotrema where they would be reflected, so that standing waves will not occur. In the direction window-helicotrema the amplitude of the waves first increases as a consequence of the decreasing stiffness of the cochlear partition; the damping, however, makes the amplitude finally decrease. The position of the *very broad*, maximum depends of the frequency, being close to the windows for high frequencies and shifting towards the helicotrema as the frequency decreases.

This localization agrees qualitatively with that proposed in HELMHOLTZ's theory of frequency analysis. Though HELMHOLTZ himself has not put it into such simple terms, his ideas are generally summarized as follows. The fibres of the basilar membrane are like the strings of a piano. The long ones close to the helicotrema resonate at low frequencies, the short stiff ones close to the windows at high frequencies. Evidently these fibres are not free to vibrate independently of their neighbours but are coupled to each other by the incompressibility of the fluid.

The experimental investigation of the cochlea is very difficult since it is a small and delicate structure. It has been especially VON BÉKÉSY (l.c.) who has succeeded in observing the motion of various cochlear structures and in measuring their mechanical constants. The agreement between experiment and theory is excellent. The broad maximum along the length of the basilar membrane for a pure tone is confirmed by these observations although according to subjective evidence our frequency discrimination is very good. Various ingenious solutions of the problem have been suggested, of which I mention especially the work of HUGGINS.[25] The theories have in common that the nervous system plays a part in them.

§ 10. *The middle ear*

The impedance at the oval window, calculated for the human ear, is still much above the impedance of air (9000/cm^2 and 40/cm^2 respectively) so that an appropriate transformer is necessary. For the lower terrestrial animals and birds this transformer consists simply of an eardrum which is connected to the stapes in the round window by a little bar, the *columella*. The soundwaves impinge on the large eardrum and the force is transmitted to the smaller oval window. Since the ratio of areas is about 1 : 10 in the pigeon, the pressure increases by a factor of 10 at the window and the allowable impedance at the stapes is now about 400 dyne-sec./cm^3. (The actual value for this bird is not known.)

In mammals, the columnella is replaced by the chain of three ossicles. These give an additional increase of allowable impedance because they reduce the amplitude by about a factor of two. Another factor of two is gained by the fact that the stapes in the oval window move not like a piston but more like a door swinging on its hinges. The *force* increases by the same total factor of four, thus making the quotient (the impedance) increase by 16. The final allowable impedance

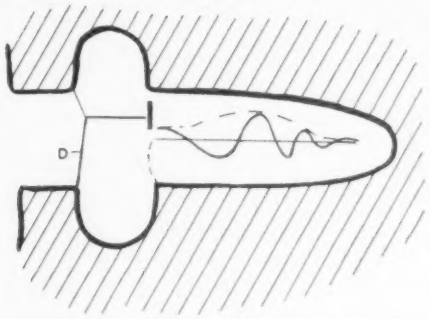


Fig. 14. Schematical drawing of cochlea and middle ear of a bird with diagram of a propagated wave of medium frequency on the basilar membrane. Note the decrease of wavelength from windows to helicotrema, by the gradual change of the constants of the canal (cross section and elasticity). The corresponding increase of amplitude is partly masked by the decrease by damping. A moment later the curve will be displaced to the right and the dotted line marks the subsequent heights of the maxima. The maximum in the dotted line depends on the frequency (see text).

at the oval window is now about $16 \times 10 \times 40 = 6400$ which is not very different from the value 9000 calculated by ZWISLOCKI (l.c.). Moreover it has been checked experimentally that the impedance of the ear, measured at the eardrum is close to 40 in the important region between about 1000 and 2000 c.p.s. (see TRÖGER[26]).

Another interesting property of the chain of ossicles is that their centres of gravity lie on their axis of rotation. This arrangement, well known in the construction of galvanometers, etc., reduces the sensitivity to shocks of the head as a whole since such shocks will not cause rotations. The sensitivity for one's own voice is also reduced in this way as well as by the coiling of the cochlea (see VON BÉKÉSY[27]).

III. The Sense Unit of the Labyrinth Organs

§ 1. *The microphonic effect*

An important tool in the study of the sense unit has been the *microphonic effect* which was discovered by WEVER and BRAY[28] in 1930. They recorded voltages from the cochlear nerve of a cat when the ear was stimulated by sound and found that voltages were a fairly good

reproduction of the wave form of the incident sound; this justifies the name microphonic effect. This experiment suggested that the cochlear nerve acted as a telephone cable, but it soon turned out that the microphonics were produced by the cochlea itself. The fibres of the acoustic nerve act as do other nerve fibres, by transmission of pulses in which the information is encoded, so that the microphonics should not be confused with the nerve signals. Perhaps a more significant anology exists between the microphonic response and the electro-

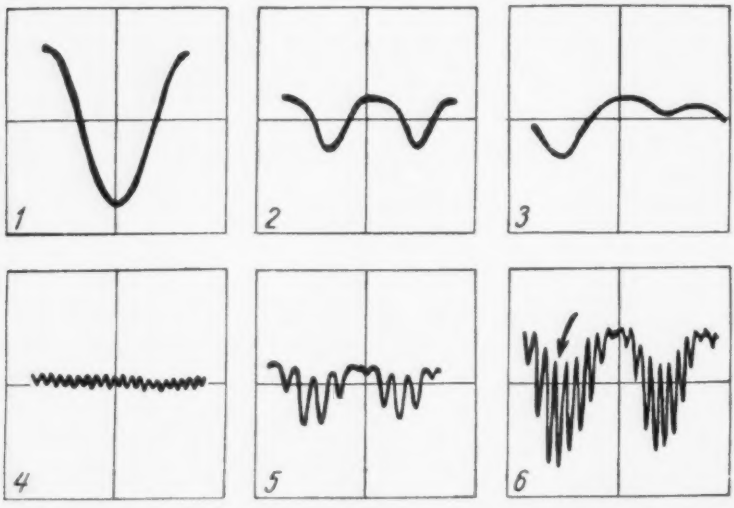


Fig. 15. Microphonic effects of the lateral line of fishes (from 32). Curve 1: Oscillogram of stimulus. Curve 2: Microphonic effect (note the double frequency). Curve 3: The same vibration, but superimposed on a displacement of the cupula. One negative peak has nearly disappeared. Curve 4: Response to a higher frequency with small amplitude. Curve 5: Response to combined stimulus of low frequency (curve 1) and higher frequency. Curve 6: the same with a higher superimposed frequency. Note the variation of the response to the high frequency component during one period of the low frequency (see text).

retinogram (the response of the retina to light) and the electroencephalogram (potentials which can be recorded from the cortex, even through the intact skull).

Since their discovery, microphonics have been studied extensively, not only for the cochlea of the cat but also for other organs of hearing of the vertebrates. PUMPHREY[29] found that even the utriculus, which has probably no acoustic function, shows a microphonic effect. From a theory of the functional significance of the microphonics, DE VRIES[30] anticipated that all labyrinth organs should produce this effect. Moreover, the scheme of Fig. 7 suggests in a simple way that it should be a

general property of all labyrinth organs, at least if it is not produced by some trivial coincidence of circumstances. Indeed the sense unit in its most simple form, the lateral line organ, produced the same microphonic effect (DE VRIES[31], [32]) and the effect was also recorded from the ampullae of the semicircular canals of the pigeon (DE VRIES and BLEEKER[33]).

The microphonics of the lateral line showed the remarkable property of having a frequency which was twice the frequency of the stimulus, i.e. the frequency of the stimulating water current (Fig. 15). This has puzzled us for a considerable time, especially since it was difficult to reconcile with our original theory (ref. [30]) according to which the microphonics were generated by a piezo-electric effect of the cupula, i.e. the production of free electric charges by deformation which is found in some crystals (e.g. those used in microphones and gramophone pick-ups). This piezo-effect is proportional to the deformation and it has always the same frequency as the deformation. There seemed to be one way to save the original theory. If the cupula is bent by the water current one side is stretched and the other compressed and voltages of opposite sign are produced. Finally one has to make the additional assumption that the structures under the cupula transmit the current more easily downwards than upwards. Such a rectifying action is not uncommon in biological structures. In this final model the two halves give in turn, a downward current in one period of the vibration and this would produce curve 2 of Fig. 15. But this theory was also disproved. First of all this model suggests that the waveform recorded at the two sides of the cupula would not be identical. Nevertheless we have always found the same waveform, even if the electrode was brought into the cupula. In another experiment nearly the whole cupula was taken away, so that only a thin layer was left. This layer is not bent but is only displaced as a whole. Even then one obtains the same output so that one has to conclude that the *displacement* of the cupula as a whole is the stimulus for the haircells (DE VRIES *et al.*[32]) and that the microphonics are not produced by the cupula. Observations of the cupula in the intact canal (in which it is placed here) by J. W. KUIPER in our laboratory showed that in this normal situation the cupula is not bent but only displaced as a whole.

These experiments have finally led to the model, given in Fig. 16. The hairs of the haircells are assumed to be embedded in little holes in the cupula, and to be connected at the top to the cupula. If the cupula glides over the epithelium, the tops of the hairs are displaced from C to C' (Fig. 16). Since OC is longer than OB they are stretched. If we assume:

1. Stretching of the hairs produces the microphonic effect, we can understand why we obtain the double frequency since the hair is

pulled again during the second half of the period of oscillation (displacement to C').

2. Pulling the hair produces a negative voltage on the epithelium at the side of the hair, we get the correct sign in agreement with the experiment (Fig. 15). Indeed we checked by stroboscopic observation of the cupula that the peaks of negative voltage corresponded to the maximum displacement of the cupula.

This model moreover explains two other phenomena. If the cupula is given a constant deviation, on which the vibration is superimposed,

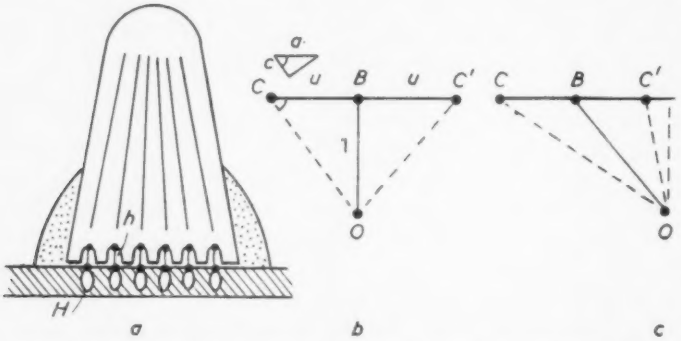


Fig. 16. (a) Schematical diagram of sense organ with haircells H and hairs h. (b) and (c) geometry of the motion of the hairs (see text).

it can give the response of Fig. 15, 3. Here one of the negative peaks has disappeared. The mechanical situation is drawn in Fig. 16c. The hair is pulled only at C. Displacement to the right (C') is not accompanied by a pull on the hair and the corresponding negative peak disappears in the microphonics.

The other effect explained by the model, is the response to two vibrations, one with a low frequency and large amplitude and one with a high frequency and a small amplitude acting at the same time. Fig. 15, 5 and 6 shows the response to the combined stimulus. The response to the high frequency is large in the negative peaks, much larger even than when the high frequency is acting alone. The response is small and it again has the double frequency when the cupula passes the equilibrium position. This latter effect is explained by Fig. 16c where one sees that the component c of the high frequency vibration increases with increasing distance u from B.

More recent quantitative analysis by J. W. KUIPER in our laboratory has revealed that some difficulties, mentioned already in ref. [32] are really of a fundamental nature. One problem is that the input-output curve (i.e. the relation between mechanical amplitude and electrical output) is linear for small amplitudes whereas the model predicts a

quadratic increase of the output (see Fig. 17b). This difficulty can be removed by the reasonable assumption that the hairs are not all exactly perpendicular to the epithelium, but inclined to the right and to the left at random. The hairs, which are inclined to the left will be pulled if the cupula goes to the left, whereas they are only relaxed for a displacement to the right. We obtain again the same waveform, but now the pull on the hairs will increase linearly with the displacement of the cupula. The explanation of the response to two frequencies is, however, now difficult. Moreover this explanation is not compatible

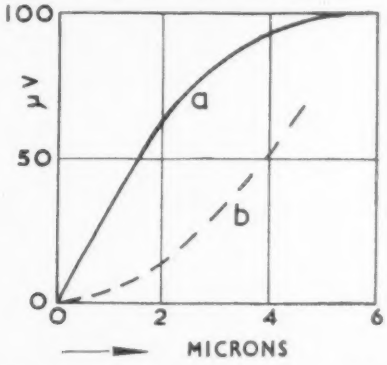


Fig. 17. Input-output curve of a lateral line organ. (*a*) Schematical, but with right order of magnitude of displacements and voltages. Curve *b*: see text.

with the tacit assumption that the output is controlled by the displacement of the hair alone. In Fig. 15, 6 the voltage nearly reaches the zero line several times in one cycle at the low frequency (see arrow). It turns out, however, that the corresponding positions of the cupula are not at all the same. Because of the small amplitude of the high frequency the cupula remains far from the equilibrium position if the displacement by the low frequency is maximum. The response must be a property of the sense cell itself and indeed the other labyrinth organs proved to give the same remarkable response to a combined stimulus (DE VRIES *et al.*[34]). For the present discussion this is not, however, of importance.

§ 2. *The functional significance and the origin of the microphonics*

The significance if the microphonics is still a matter of hypothesis, but the most reasonable assumption is that the branches of the nerve fibres which surround the sense cell are stimulated by the microphonic currents which flow from the lower side (which is positive during stimulation) to the upper side (the side of the cupula). Indeed one can also get a response of a labyrinth organ—i.e. the same effect and sensations as are produced by adequate mechanical stimulus—by

electric stimulation. To achieve this response the negative electrode has to be above the organ, which means that the current has the same direction as the "microphonic" current.

For a discussion of the origin of the microphonics it is important to know that the energy connected with the microphonic currents is larger than the energy of the mechanical stimulus which displaces the cupula. For the organ of the lateral line our proof of this was not

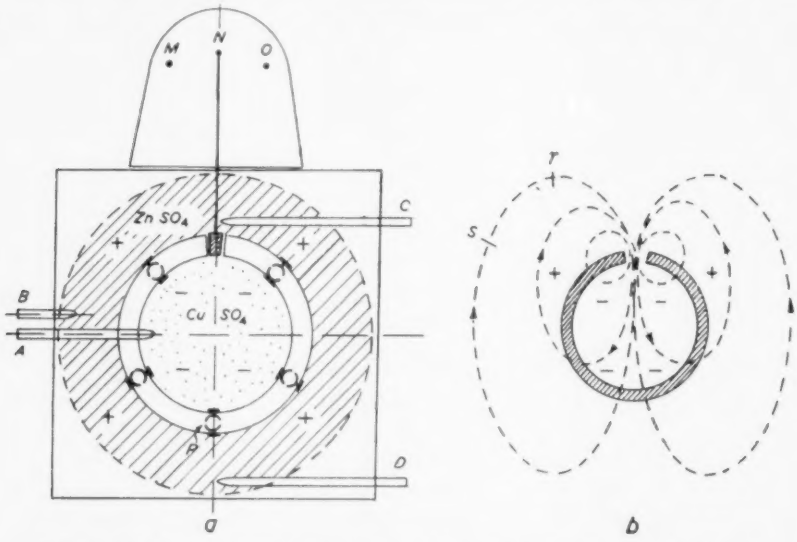


Fig. 18. Model of a haircell, with the hair connected to the cupula in *N*. By displacement of the cupula the conical stop is pulled out of the cell wall so that the cell is more or less polarized; (*b*) gives (schematically) the electric currents. Relative to the "indifferent" electrode *B*, *C* becomes negative, whereas *D* becomes positive.

unambiguous (DE VRIES, l.c.) The energy dissipated in the media around the organ was nearly the same as the mechanical energy but it was probable that much more electrical energy was dissipated in the organ itself, thus making the unbalance between electrical and mechanical energy considerable. In similar measurements in the cochlea VON BÉKÉSY[35], was able to prove, however, that the electrical energy exceeded the energy of the mechanical stimulus.

The data assembled so far on the origin of the microphonics can be combined in the following model which we have made of the sense cell with its hair (Fig. 18). The inside of the cell is assumed to be negatively charged relative to the outside by, say, 100 mV, which is quite normal for living cells. This polarization is produced and kept constant by metabolic processes. If the hair is pulled, the isolation and polarization of the cell wall are assumed to be reduced at that place. A current will

flow into the cell, producing a negative voltage at the side of the hair. At the same time electricity is pumped out by the metabolic processes in the rest of the cell wall. The actual model, which we have made of plexiglass, has two spaces filled with fluid, which are separated by the cell wall, an empty ring. The polarization is produced by voltaic cells, consisting of copper electrodes *in* the cell, which contains copper sulphate and zinc electrodes in the medium which contains zinc sulphate. They can be connected and disconnected by means of stoppers P in the cell wall. Electrodes A, B, etc., measure the distribution of voltages

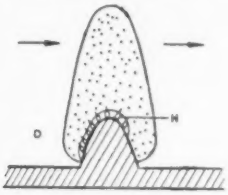


Fig. 19. Schematical drawing of the crista in the semicircular canal. Left-hand side D: dominating side. H: haircells. Direction of arrow: "stimulating" direction.

around the cell. A cone, which can be withdrawn by a wire, closes the hole between the medium and the inside of the cell.

§ 3. *The sense unit in the other labyrinth organs*

We have recorded the microphonic effect of the sacculus of the perch; it gave the double frequency (DE VRIES *et al.*[32]). For the lagena otolith, VON HOLST[36] has shown that the displacements of the otoliths parallel to the epithelium were the active stimulus, and not pressure on the epithelium, etc. Previously there have been many contradictory views on this subject. If one considers the labyrinth organs as one group, it is evident that only forces parallel to the epithelium could have functional significance. Of course this does not decide whether deformation or displacement of the cupula is of primary importance.

The microphonic effect of the semicircular canals has the same frequency as the mechanical vibration. There is histological evidence, however (VILSTRUP[37]), that the hairs are inclined (see Fig. 19) and we have already seen (Fig. 15, 3) that for this situation the output has the single frequency. It should be noted that on both sides of the crista the inclination is in opposite directions but one side of the epithelium dominates (in Fig. 19 the left side), consequently a deviation in the direction of the arrow will produce a much larger effect than a displacement in the opposite direction. The physiological corollary of this assymmetry is that a deviation in the direction of the arrow stimulates the organ whereas a displacement in the other direction does not

(EWALD's law). For each of the three canals the relation between the stimulating direction and the dominating side is the same as described above. VILSTRUP (l.c.) has also shown that the cupula is not deformed much, but is displaced when a force acts on it.

Finally, in the cochlea, the situation is very complicated. The single frequency observed is easily explained by the inclined position of the sense cells at each side of the tunnel of CORTI. VON BÉKÉSY[35] has shown that in the resting organ d.c. potentials can be recorded. Probably the sign is in accord with the model of Fig. 18 and the assumption is that these potentials correspond to a constant "leak" along the cone which closes the hole. In other experiments he showed that the shearing motion of the tectorial membrane over the organ of Corti produces the microphonics. This is in accord with our findings for the more simple lateral line organ.

In all cases where a check was possible the sign of the voltages was in accord with the rule that pulling the hairs produces a negative voltage above the organ (and positive beneath it) (DE VRIES and VROLIJK[38]).

IV. BROWNIAN MOTION AND THE LABYRINTH ORGANS

§ 1. *Introduction*

During this century the fundamental limits which are set to the performance of physical instruments have been gradually understood. At the same time the sensitivity has been forced up to these limits. Undoubtedly the interest of physicists in biology has been stimulated by the discovery that in the sense organs nature has already reached these same limits long ago.

In the case of light perception and vision, the limits are due to the quantal character of the stimulus itself (see next section). In other cases the limits finally arise from the thermal vibrations of the molecules of which the organs are composed. This is most directly true for the detection of differences in temperature.

Not very much is known about the living thermometer, but there must be somewhere a region, containing N atoms, the temperature of which is detected. We can consider the atoms as being embedded in a solid structure; consequently each molecule has 6 degrees of freedom and an average energy of $3kT$. If we make an instantaneous photograph of this volume, however, the total energy will not be exactly $6N$ times $\frac{1}{2}kT$ since fluctuations of the order of $\sqrt{6N}$ will occur. This means that the temperature of this small volume is no longer T, but that it fluctuates by an amount ΔT, for which $\Delta T/T = \sqrt{6N}/6N$, since the temperature is proportional to the energy. If the organism is at 27°C, T is equal to 300 and $\Delta T = 300(6N)^{-\frac{1}{2}}$. In a spherical region with a diameter of 1 micron (a small part of a cell of medium size) N equals

about 10^{11} and $\Delta T = 4 \cdot 10^{-4}$°C. Though deviations of this magnitude are fairly frequent, it is obvious that they are very small. If the detecting unit is even smaller, however, the fluctuations become more disturbing.

One other aspect has to be considered. We considered only the situation in one "snapshot." If the average life time of one "snapshot" is, say, one millisecond, a deviating temperature will be less disturbing for the measurements than if each temperature configuration lasts for several seconds. This average life time of a configuration is important in all fluctuation problems. In our case it can be calculated from the thermal conductivity of the medium and its heat capacity. It depends on the volume considered and in our case we find for the average life time of a deviation about 10^{-5} sec. If the detector could be triggered immediately if the temperature exceeded a certain value, we would consequently have 10^5 chances per second and the threshold of the detector should be high enough so that in less than about one in 10^5 of the snapshots this threshold is attained. The other extreme case is that of a detector triggered by the average energy in a time of integration T. If this time contains n life times of one configuration the fluctuations are averaged and the final fluctuations in the average value are reduced by a factor of $\sqrt{n}$.

In the following sections we shall confine ourselves to Brownian motion in the ear. A thesis on the other labyrinth organs is being prepared in our laboratory. SCHMALTZ[39] has already calculated the Brownian motion in the semicircular canals, but before it was recognized that the elasticity of the cupula should be taken into account, and that the cupula was displaced by the displacements of the endolymphe.

§ 2. *Brownian motion and hearing*

The greater part of this section is a summary of previous publications (DE VRIES[30], [40]). The Brownian motion will be considered at three stages in the processes by which sound is finally "transformed into" nerve discharges. We start with the Brownian motion of the air at the eardrum. After that we consider the Brownian motion of the eardrum itself and finally the Brownian motion at the sense cells.

Brownian motion of the air. The eardrum is bombarded from both sides by molecules. Macroscopically this bombardment gives rise to the pressure of the gas, which, on the average, is the same at both sides of the eardrum. Statistical fluctuations will, however, occur and these can be interpreted as Brownian soundwaves. If the average number of collisions in a certain time (which depends on the bandwidth of frequencies considered) is equal to n, the fluctuations will be of the order of $\sqrt{n}$ and this can be converted into the corresponding fluctuation in pressure. Since n is easily calculated one would think that the Brownian

soundwaves are easily found. Indeed, this method gives the right order of magnitude and a good idea of their origin. We have tacitly assumed, however, that the various collisions were independent of each other, whereas, in reality, a correlation exists over a distance which has a range of about one wavelength λ of the soundwaves considered. One obtains the true value, if the number of collisions n is considered over an area as large as about $\pi\lambda^2$. If the ear is provided with a tube with the same cross section as the eardrum the smaller area of the eardrum has to be inserted (equation (5)).

In a straightforward way the amplitude of the soundwaves is obtained from the following model. The eardrum is represented by a membrane in an infinite plane. It is connected to an electrodynamic loudspeaker circuit (which can act at the same time as a microphone). Soundwaves will produce a dissipation of energy in the resistance R that closes the circuit of the coil of the loudspeaker. In a similar way currents in R will produce soundwaves. According to the "rule of detailed balance" as much energy should flow out in a certain region of frequencies as flows in. Now the outward flow can be calculated since we know the spontaneous e.m.f. in the resistance which is given by the formula of NYQUIST

$$\langle V^2 \rangle = 10^{-7} \cdot 4kTR\Delta\nu \qquad \ldots.(1)$$

Here V is measured in volts, kT in ergs and R in ohms. The expression for the inward flow contains the unknown pressure of the Brownian soundwaves which can now be calculated. The equality of inward and outward flow gives the equation

$$4\langle p^2 \rangle F^2 w_e/(w_e + w_a)^2 = 4kT\Delta\nu w_e w_a/(w_e + w_a)^2 \qquad \ldots.(2)$$

where $\langle\rangle$ means "average value," w_e and w_a stand for the impedance (quotient of force and velocity amplitude) of the ear itself, measured at the eardrum, and of the *radiative* resistance of the ear, i.e. the impedance due to the fact that energy is dissipated as soundwaves if the eardrum vibrates; F is the area of the eardrum. Obviously details of the construction of the loudspeaker do not go into the formula, only the impedance w_e is of importance, and even this cancels out if we are only interested in the Brownian soundwaves, since w_e appears in the same way on both sides of equation (2). The resistance w_a can be found from textbooks on acoustics:

$$w_a = 2\pi\rho c F^2/\lambda^2 \text{ dyne/cm/sec.} \qquad \ldots.(3)$$

Here ρ stands for the specific weight of the air (λ being the wavelength and c the velocity of sound). The formula holds for $F/\lambda \ll 1$. We are mainly interested in frequencies below, say, 3000/sec., and indeed for these frequencies F/λ is smaller than 1. Moreover (3) gives only the

real part of w_a, since we only need that part. Inserting this value of w_a in (2) one obtains the expression for the Brownian soundwaves which has already been given by SIVIAN and WHITE.[41] In their notation p is the pressure at the eardrum which is equal to $2p$ in our notation.

Though w_e cancelled out in the foregoing calculations, it will be required in the following section. According to various measurements (see also § II, 9) w_e is nearly real in the most important region of frequencies, between about 500 and 3000, and equal to the impedance of the air (ρc per cm^2). Since F is about 0·3 cm^2 w_e is 12 c.g.s. units. w_a is only about 0·05 at 1500/sec. For comparison with the sensitivity of the ear it is more interesting to know the apparent flow of energy into the ear, than the amplitude of the Brownian soundwaves. For the "free ear, where w_a is given by (3), one finds from (2)

$$I = kT(8\pi F/\lambda^2)\Delta\nu \qquad \ldots.(4)$$

At 1500 c.p.s. the coefficient of kT in (4) is about 0·02. If the ear is provided with an ear trumpet (exponential horn) the radiative resistance w_a increases considerably becoming ρcF, (equal to w_e). The apparent inward flow of energy now becomes

$$I = kT\Delta\nu \qquad \ldots.(5)$$

Though (4) and (5) give us the inward flow of energy from the air, it is of still more importance to consider the vibrations of the eardrum itself, since they can tell us whether anything will be heard or not. A velocity-amplitude u of the eardrum corresponds to a flow of energy of u^2w_e. If u_e and u_a stand for the velocities due to the ear and to the air, respectively we have $\langle u^2\rangle = \langle u_e^2\rangle + \langle u_a^2\rangle$ since they are independent of each other. The contributions u_e^2 and u_a^2 follow from (2) since the two sides give the flow of energy $u_a^2w_e$ and $u_e^2w_a$, respectively. So one finds for the free ear u^2w_e:

$$I = 4kT\Delta\nu \qquad \ldots.(6)$$

If an ear trumpet is used:

$$I = 2kT\Delta\nu \qquad \ldots.(7)$$

A comparison of (6) and (7) demonstrates that the effect of the ear trumpet is quite different from what is suggested by a comparison of (4) and (5). Though the trumpet gives an increased exchange of energy between the ear and the soundfield, the amplitude of the eardrum is decreased by the stronger damping.

§ 3. *Comparison of the Brownian motion of the eardrum and the absolute threshold of the ear*

Before we can answer the question whether the Brownian noise is above or below the absolute threshold of the ear, we want to know what this

threshold is, and also the values of ν and $\Delta\nu$ that have to be inserted in the formulae. A reliable value for the threshold of a good observer is about 20.10^{-12} erg/sec. at 1500/sec. (for a discussion of the literature see DE VRIES[42]). For ν we shall insert 1500/sec., since the ear has a maximum sensitivity around this frequency. What value to insert for $\Delta\nu$ is less obvious. SIVIAN and WHITE[41] took for this bandwidth the whole region between 1000 and 6000/sec. It is doubtful, however, whether the excitations by the various frequencies of this band would be simply additive. It is more probable that one should take a band such that the regions of maximum stimulation on the basilar membrane overlap (see also Fig. 14). This bandwidth is 300/sec. around 1500/sec. One finally finds from formulae (4)–(7) for I the values 0·2; 12; 48 and 24 (10^{-12} erg/sec.). According to (4) and (5) the Brownian motion of the air, even with an ear trumpet, is below the threshold. It has already been emphasized above, however, that one should consider the amplitude of the eardrum and disregard the origin of the vibrations. In this case we see that for a free ear (form (6)) the noise is above the threshold. It should be kept in mind, however, that this threshold is determined in the presence of the Brownian noise. It is possible that our result has to be interpreted as follows. One can assume that the actual threshold is 68.10^{-12} erg/sec. The Brownian noise supplies 48.10^{-12} and the "threshold-signal" has to supply 20.10^{-12} erg/sec. Moreover, the thermal noise is not constant; it will reach the supposed threshold only very rarely, however (see next sections).

According to (6) and (7) the noise decreases if the ear is provided with an exponential horn (a long rubber tube gives the same acoustical impedance). We have done this experiment but we did not find an "audible" effect. It is still of interest to measure the absolute threshold in the two cases in order to check the difference that is suggested by the preceding paragraph.

§ 4. *General discussion of the Brownian motion in the sense cells*

In the preceding sections the Brownian motion at the "entrance" of the organs was considered. Now we shall study the effect of thermal motion on the sense cells themselves. Although a complete discussion is only possible if we know the processes in the sense cell, some general remarks can be made which are of importance for any sense organ.

First of all it is of importance that the nerve fibres transmit the information encoded in pulses of standard size and with a maximum repetition rate of about 500/sec. We do not know whether the sense cells or at least the cells of some sense organs, respond in the all-or-none way which is typical for the nerve fibres. At any rate we can represent the cell plus its nerve fibre by a *gun*. A shot goes off when the displacement of the trigger exceeds some minimum value.

VOL
6
195

Most important for the Brownian motion is the way in which the "main trigger" (the source of the stimulus) is coupled with the triggers of the individual guns. We shall consider two extreme cases. First what can be called "hydraulic" coupling. The "main trigger" moves a large piston in a cylinder; the pressure is transmitted to smaller pistons which are coupled to the triggers of the individual cells. In our second model the coupling between main trigger and cells is supposed to be a single rigid bar. In several respects the two systems are equivalent, especially when the area of the large piston is equal to the sum of the areas of the small ones. In this case the same force, the same displacement and the same energy are necessary in both models in order to get at least one shot.

The hydraulic coupling gives more Brownian motion, however, since any of the small triggers can move almost independently of the others. The probability of a spurious shot mainly depends on the energy needed for firing the gun (see below) and this energy is obviously n times smaller in the case of hydraulic coupling than in the case of a rigid coupling of n guns.

The two systems considered are only two extreme cases. When, for instance, the triggers are connected to the rigid bar by springs, then they can also move independently of each other, just as in the case of hydraulic coupling. Finally, leaks in the hydraulic coupling, or slip in the rigid coupling, will make things much worse, since the Brownian motion in the cell will still be present or perhaps even increase, whereas more energy is required at the input.

To conclude, it is of importance to estimate more quantitatively the number of spurious excitations. In order to fire a gun, a certain minimum energy E has to be applied to the trigger. The probability that this energy (or more) is concentrated on the trigger by Brownian motion at some time is given by the Boltzmann factor $\exp(-E/kT)$. After a correlation time t_0, which depends on the system, a different energy will be concentrated on the trigger. When the number of such "new" energy configurations is f per second ($f = 1/t_0$) the number of spurious excitations per second will be $f \exp(-E/kT)$. In the case discussed in the first section of this chapter (heat receptor), f was 10^5, and in several other cases f turns out to be much larger than unity (see DE VRIES[43] and also the following sections). As a result, E has to be much larger than kT in order to avoid a large number of spurious excitations. We can add here that the stimulus need not provide this threshold energy E completely, since the Brownian motion, which does not attain the threshold energy alone, can help to produce the necessary stimulus to trigger the cell. This was already anticipated in the foregoing section, but it will be discussed in more detail in section V, 7 (visibility of red light).

A system can also act in a more complicated way than does a gun. Instead of a system reacting to the instantaneous value of the displacement of the trigger, one can imagine an electric relay containing a rectifier, combined with a condenser which will integrate the voltage over a certain time, depending on the leak of the condenser. When the time of integration T is m times the correlation time t_0 defined above, the output of the integrated noise will, on the average, be mkT, with fluctuations Δ which are of the order of $kT(m)^{\frac{1}{2}}$. Since only the fluctuations in the noise are disturbing, the effect of the integration is that for the same time interval the disturbing effect is $m^{\frac{1}{2}}$ times smaller. If we gave the relay a threshold of $mkT + 2\Delta$, spurious excitations would be fairly rare, whereas a real signal of about 2Δ would be sufficient to get a response.

Up to now we have tacitly assumed that the number of spurious excitations should not be too large. Indeed, if there were many such excitations it would mean that the whole system of nerve fibres would be overloaded by Brownian noise, and would not be available for real signals. A small number of spurious excitations need not, however, be very disturbing. In this latter case there is the possibility of discriminating between real pulses and spurious pulses since the real excitations tend to be synchronous, whereas spurious pulses are not. This is obvious in the case of the labyrinth organs where the real pulses arise from displacements of the cupula, which are synchronous for all the haircells. Direct evidence for such a "coincidence system" has been demonstrated for the eye (see § V, 6).

§ 5. *Brownian motion in the inner ear*

We shall now apply the results of the foregoing section to the inner ear. In section III, 1, evidence was given that the cells were stimulated by pulling their hairs. The hairs are embedded in the cupula (tectorial membrane) and if the coupling is rigid enough, the whole system will be identical with our model of triggers, coupled by a rigid bar. Under these conditions the signal-noise ratio at any hair would be the same as at the eardrum, and this means that the signal would be stronger than the noise. It will now be shown, however, that the coupling is not rigid enough.

It is of interest to estimate first the stiffness required for a good transport of energy to the sense cells without considering Brownian motion. From the present knowledge of the mechanics of the middle ear and of the cochlea (see section II, 9) one can very roughly estimate that the amplitude of the hairs is of the same order of magnitude as the amplitude of the eardrum. Since the amplitudes are equal, the impedance of the eardrum must be the same as the impedance of the 3000 haircells which are stimulated; that it is 0·004 c.g.s. units per

hair (for the number of 3000 cells, see below). This means that 0·004 dyne gives the base of the hair a *velocity* of 1 cm/sec. The *linear* amplitude of the tectorial membrane is then $1/2\pi\nu$, where ν is the frequency so that the stiffness f of the hair should be sufficiently large for the hair to be stretched less than $1/2\pi\nu$ by a force of 0·004 dyne. At 1500/sec. we find that f should be at least 38 dyne/cm.

The actual stiffness can be estimated as follows. Assuming that the hair itself is fairly rigid and embedded in the softer material of the tectorial membrane, we have the situation of a hair embedded in a pudding. The stiffness by which such a hair is held is $f = 2\pi lG/\ln r_e/r_i$ dyne/cm where l is the length of the hair $= 10\,\mu$, $G =$ elastic constant of the tectorial membrane $= 10^4$ c.g.s. units, $r_i =$ radius of the hair $= 0{\cdot}5\,\mu$. r_e is the radius of the cylinder in which the hair is supposed to be embedded and for this we can take the distance of two cells—about $10\,\mu$. The numerical constants have been taken from unpublished measurements on lateral line organs, partly described in section III, 1. From the formula we find $f = 20$ dyne/cm. Because of the uncertainties in some of the numerical values the agreement with the "required" value estimated above is more than good.

Coming now to the discussion of the Brownian motion, we shall first prove that the stiffness calculated above is not great enough to give the rigid coupling defined in the foregoing section. When the hair is displaced over a distance d, the potential energy equals $\frac{1}{2}fd^2$. For the threshold (as assumed in § 3) of 20.10^{-12} erg/sec. the displacement measured at the eardrum is $1\,.\,4{\cdot}10^{-10}$ cm at 1500/sec. Since the amplitude d of the hairs is roughly the same, the energy of the stretched hairs is $\frac{1}{2}fd^2 = 4{\cdot}10^{-19}$ erg. This is so much less than $kT(4{\cdot}10^{-14}$ erg) that the sense cells can be considered as completely free. The difference is a factor of 10^5 and this is much more than can be accounted for by the possible errors in the constants inserted, we have therefore to consider the Brownian motion of the individual cells.

For this discussion the absolute threshold of the ear as a whole is taken as 20.10^{-12} erg/sec. $= 500\,kT$/sec. In short signals (up to a few tenth's of a second) the threshold is higher if the *intensity* (erg/sec.) is considered. Here the total energy is constant, independent of the duration of the signal, and equal to $0{\cdot}4 \times 500 = 200\,kT$. The time of 0·4 sec. is the effective time of observation. (See DE VRIES[42] and a recent survey by NEWMAN.[44])

According to our present knowledge of the mechanics of the cochlea about 5 mm at least of the basilar membrane is vibrating, even when the ear is stimulated by a pure tone. This region contains at least 3000 sense cells.

First we shall prove that the sense cells cannot merely act as triggers. The damping time t_0 of the cochlea is about 3 millisec. When the

stream of energy is $500\,kT$/sec., during any time interval t_0 only $0{\cdot}003 \times 500 = 1{\cdot}5\,kT$ is available. This, moreover, has to be distributed over 3000 cells, all of which already have $\frac{1}{2}kT$ per degree of freedom. As a consequence far too many spurious excitations would be triggered by thermal motion if the cells were activated by the instantaneous value of the displacement of the velocity. The ear, at least at threshold, must have the property of integrating over longer periods before it comes to all-or-none processes.

Even when the ear integrates the sound over 0·4 sec. (the effective time of observation at threshold) there remains a discrepancy. The energy per 0·4 sec. is $200\,kT$. It is distributed over 3000 sense cells so that one cell gets $0{\cdot}06\,kT$. The bandwidth of one cell, as shown by measurements on the masking of tones that differ a small amount in frequency, is at least 300/sec. The Brownian noise therefore corresponds to $300\,kT$/sec or $120\,kT$ in 0·4 sec. According to the preceding section the fluctuations will be about $(0{\cdot}4/0{\cdot}003)^{\frac{1}{2}}\,kT = 11\,kT$, so that the signals per cell, even after integration, will be still smaller than the noise.

Needless to say, these discrepancies would become even more serious when only part of the energy were available for the sense cells. One solution to the difficulties was suggested by DE VRIES.[30] The fundamental suggestion was that more than a proportional part of the energy was concentrated on a small number of sensitive cells, so that for these cells the signal would be larger than the noise. One easy way of achieving this would be by transforming the incident mechanical energy into electrical energy which could then pass to the sense cells with the lowest electrical resistance. The most appropriate place for this transformation would be the tectorial membrane, the voltages generated being the cochlear microphonics. If the microphonics really have this fundamental significance, they should be found in all labyrinth organs and in the lateral line, and in section III, 3 it has been shown that this is so. But further analysis of the response of the lateral line showed that the microphonics were not generated by the tectorial membrane, but by the sense cells as a response to the pulling of the hairs.

The new model developed in section III, 2 can, however, solve the discrepancy. The recorded potentials and currents are the summed response of a large number of cells. The response to a real stimulus has a constant phase relation for all these receptors, whereas the Brownian currents have no special phase relation and therefore partly cancel. The increase of the current resulting from Brownian motion is thus proportional to the square root of the number of cells, whereas for a real stimulus the response is proportional to the number of cells (at least for parts of the basilar membrane which vibrate in the same phase). The improvement in the signal-noise ratio is therefore proportional to the square root of the number of co-operating cells. We do not

know how many cells this is, but if it is the number in a length corresponding to one-half wavelength of the propagated wave in the cochlea, then for a 1500 cycle tone at the position of maximum amplitude we find that it is about 3000 cells. The improvement would thus be by a factor of 55. The Brownian fluctuations would correspondingly be reduced by a factor of 55 relative to the real signal. The signal noise ratio, which was 0·06/11 now becomes 0·275. This is still below 1, but the difference may be due to errors in some of the constants inserted, especially the value of the *minimum audibility*. Once the signal-noise ratio in the microphonics is high enough, subsequent processes leading to stimulation of the nerve fibres need not give trouble since the microphonics contain much more energy than the incident sound (§ III, 2). Consequently the energy of thermal disturbances etc. in these subsequent processes are much below the level of the energy in the microphonic effect. It is always the first stage of the amplifier which gives the noise (or the signal itself). Moreover (at least at threshold) the integration over times of about 0·4 sec. finds its place in the model. The net effect of pulling the hairs is always a current that depolarizes the cell. or a current flowing from the base of the epithelium to its upper side. If the hair vibrates the current is not constant, but this is not essential for the present purpose. The only additional assumption that has to be made is that the nerve branches in the epithelium respond to the integrated current of the last 0·4 sec. Such an integration is not uncommon for the electrical stimulation of nerves; only the time of integration is often shorter.

VOL.
6
956

V. Physical Aspects of Vision

§ 1. *Introduction*

In the same way as optics is subdivided into geometrical optics, wave optics and the quantum theory of light, the physical aspects of vision can be considered from these points of view. Geometrical optics deals mainly with the dioptric apparatus of the eye: since there have been no important new developments, we shall not discuss it further here. The main consequence of the wave character of light is that it limits the visual acuity of the eye; this will be discussed for the insect eye, together with its ability to detect the plane of polarization of polarized light (another aspect of the wave-character of light). Finally the directional sensitivity of the sense cells belongs to this chapter since it is the counterpart of the directional effect of dielectric antennae.

Perhaps most interesting from the point of view of modern physics are the quantum effects in the visual process. In a previous volume of *Progress in Biophysics*, Pirenne[45] has given a survey of this part of our subject, so that we confine ourselves to some special topics. Finally

the chapter will be concluded with a discussion of effects of thermal agitation on the visual process.

§ 2. *Effects of diffraction for one single sense cell with a lens**

Fig. 20 gives schematically the arrangement of one sense cell with a lens of diameter d. Even for an ideal optical system and for a very small diaphragm D in front of the cell the system will not only be sensitive to light rays parallel to the optical axis but light coming from other directions can also contribute to the illumination of the cell by diffraction at the entrance pupil. For the present purpose we can reverse the direction of the light rays. The beam of light will be diffracted so that

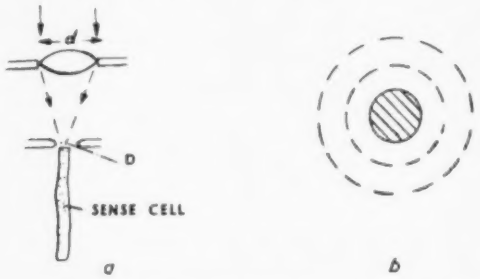


Fig. 20. (*a*) Sense cell with lens. (*b*) Airy pattern, giving in this case the distribution of directional sensitivity by diffraction.

the distribution of the directional sensitivity of the system in a plane at some distance will be given by the well-known AIRY pattern (Fig. 20*b*).

Since the intensity in the central maximum is much greater than in the rings around the centre we shall confine ourselves to this centre. The cone which has its apex at the lens and the central Airy disc as its base has a top-angle

$$\phi = 2 \,.\, 1{\cdot}24 \,.\, \lambda/d \text{ rad.} \qquad \ldots.(1)$$

(For diffraction at a slit with width d the angle ϕ would have been given by the same formula, but without the factor 1·24.) We can say that this angle ϕ is approximately the minimum angular distance that can be "resolved" by the system. It is of interest, however, to discuss this statement in more detail.

As the pattern which has to be resolved we have chosen in Fig. 21 a grating, consisting of white and black bands. In Fig. 21*a* the axis of the eye passes through a black band whereas in *b* it passes through a white band. It is evident that the pattern can be resolved only if

* *Note added in proof.* Recently I found that BARLOW[75] gave a discussion of diffraction in insect eyes that differs from the present discussion mainly where it does not take into account the fact that one partial eye sees seven or eight points in the visual field instead of one. (See also ref.[46].)

situations a and b produce illuminations which can be discriminated by the cell. For one cell, or at least a small group of cells, the minimum perceptible difference in intensity is of the order of 10 per cent (see DE VRIES[47]) for some measurements on the frog eye; for the insect eye we have no data). If the bands are very wide, compared with the Airy disc in this plane, situation a will produce no illumination of the cell, whereas b will give maximum illumination. For a very narrow grating the two situations will produce equal degrees of illumination, which are one-half of the maximum. Obviously this pattern cannot be resolved. Making the wide grating gradually narrower, we calculated by graphical integration that the width chosen in Fig. 21 gives a

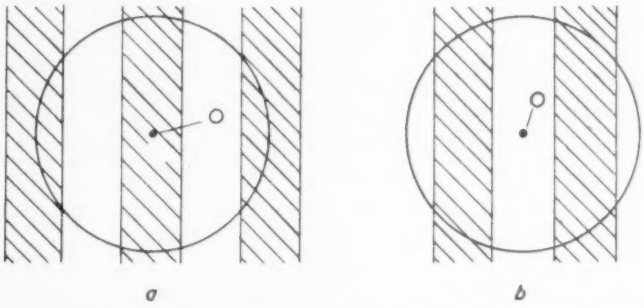


Fig. 21. Black-and-white grating in the visual field (centre O) of one sense cell. (a) O on a blackstrip; (b) O on a white strip.

difference of 10 per cent between a and b, so that it is obviously just the minimum perceptible width. This difference of 10 per cent falls off very rapidly with decreasing width, so that the choice of 10 per cent is not essential. For a somewhat narrower pattern the difference between a and b even becomes negative—this means that a gives a greater brightness than b.

If w is the angular distance of two lines of the grating, it follows from the foregoing discussion that w should be at least $\frac{1}{2}\phi$, or

$$w \geqq 1{\cdot}24\lambda/d \qquad \ldots.(2)$$

Up to now we have assumed that the diaphragm in front of the receptor was very small. If this were so, however, the receptor would hardly receive any light. On the other hand the diaphragm cannot be too large since this would imply a very bad visual acuity. Generally the best compromise in such situations is to make the contribution of the two sources of error equal. In our case this means that the diaphragm D, as seen from the lens, should subtend the same angle as the Airy-disc (ϕ). If the space between lens and cell is filled with material of refractive index n, this angle ϕ should be n times smaller.

The angle w, defined above, increases of course with increasing D, but not completely by a factor of 2 for the chosen value of D, since the intensity of the light falls off to the edges of the visual field of the cell. For a Gaussian distribution the minimum perceptible angle w increases by a factor of $\sqrt{2}$ and we shall assume a factor of 1·7 (the average of 2 and the square root of 2). The exact value is not of interest here. Denoting this value of w by w_t, we have

$$w_t = 2{\cdot}1\ \lambda/d \qquad \ldots.(3)$$

In the preceding paragraph we have assumed that the diameter D of the cell (or its diaphragm) subtends an angle ϕ at the lens. If the focal length of the lens is f, then:

$$nD = 2{\cdot}48\ \lambda f/d \qquad \ldots.(4)$$

The quotient f/d is the "stop" of the lens. f/d is not the same for all insect eyes but we can use here an average value of 3. For the human eye with wide pupil f/d is also approximately 3. The stop number governs the brightness of the image produced by the lens. In this respect there is obviously no essential difference between insect eyes and vertebrate eyes. Formula (4) suggests that the optimal diameter D of the cells would also be the same for insects and mammals, and would be about 6 λ, or about 3 μ. For vertebrates the diameter of the rods is about 3 μ; for insect eyes (calliphora) the diameter is not known exactly, since we do not know the diameter D at the top of the cell. Our impression is that it is somewhat less than 2 μ; inserting $\lambda = 0{\cdot}4\ \mu$ for the insect eye we calculate from (4) $D = 2\ \mu$.

§ 3. *The optimal construction of a compound eye*

In the preceding section we found some points of similarity between eyes of insects and of vertebrates. There is, however, one essential difference in construction. The vertebrate eye contains one lens, which gives an image on one retina, whereas the insect eye is compound, containing a large number of partial eyes, each with one lens and its own retina. (Apart from these "normal" eyes there is also a median eye, the *ocelli*, which will not be considered here.) The "retina" of each partial eye requires special discussion. In the greater number of insects it contains 7 or 8 separate sense cells; in certain groups (the *hymenoptera*, e.g. the bee) these cells together form one cylinder along the axis of which the rhabdomers form together one central rhabdom (see also § 5). Since Exner's work (ref. [48]) on the insect eye it has been generally assumed that even in the case of separated sense cells the receptors do not "see" different parts of the image. Exner does not describe how he arrived at this conclusion, even though he originally held the opposite view that the separate sense cells see different parts

of the image, just as in the vertebrate eye*. According to some unpublished measurements (see ref. [49] and § 5) this latter view is correct.

We shall consider now the question whether the compound eye or the vertebrate eye gives the better results. We shall consider both eyes as spheres of the same radius R. Taking the diameter of the lens of one partial eye as d, the angle a between the optic axis of adjacent eyes (disregarding some irregularities), will be $a = d/R$. The number of ommatidia along an arc of 1 radian will be R/d. There are 8 times as many sense cells as partial eyes, therefore along an arc of the sphere there will be $\sqrt{8} \approx 3$ times as many cells as partial eyes. If this insect eye looks at a pattern consisting of a grating of black and white strips it is clear that in order to make it possible to "see" the strips of the grating, i.e. to detect the difference between a more or less evenly illuminated plane and the grating, the number of black strips included in an angle of one radian must not be larger than one-half of the number of detecting cells included in this angle. If the number of cells is twice the number of the black strips, adjacent cells will see successively a white and a dark strip. If the visual angle between adjacent black strips is w, we now have

$$w \geqq 2d/3R \qquad \ldots.(5)$$

Combining this with (2) we have

$$d^2 = 3{\cdot}2\,\lambda R \qquad \ldots.(6)$$

and

$$w^2 = 1{\cdot}4\,\lambda/R \qquad \ldots.(7)$$

For the eye of calliphora we can insert for R about $750\,\mu$. For λ we choose $0{\cdot}45\,\mu$, which is somewhat to the blue side of the maximum sensitivity of the human, a shift which is indicated by various experiments. Calculation gives $d = 33\,\mu$ whereas the observed value for calliphora varies from 20–40 μ. The visual acuity w calculated from (7) is $1{\cdot}5°$. It is generally stated that w is approximately equal to the angle between adjacent ommatidia, d/R rad $\approx 2{\cdot}5°$, but smaller values have been reported. The fairly good agreement between calculated and observed suggests that our starting point was right. One of our assumptions, however, is somewhat doubtful. We assumed that each receptor only perceives one element of visual space. Yet if the insect *moved* its eye it could analyse the whole visual space with one receptor only. It could do so by turning its whole head, or during flight. It has been emphasized by AUTRUM[50] that the rapid motion of an image over the cells of the retina, which cannot in an insect be compensated by rotation of an eyeball, has lead to the development of a "temporal"

* Perhaps a discussion of diffraction in the insect eye by MALLOCK in 1894 (see ref. [48]) had some influence.

analysis of the image by the insect, rather than a spatial analysis. This suggestion is supported by the very high critical fusion frequency observed for insects—i.e. the highest frequency that does not give the impression of continuous illumination. For man it is about 20 (depending on brightness, etc.) whereas for the insect eye it is nearly 300. It should be kept in mind that the scanning of visual space by too small a number of receptors would only be effective for a static pattern. Since moving objects are of especial biological interest, it is important to "see" the whole visual field in one glance, which, finally, was our starting point on the previous page.

Finally we compare these results with the performance of the vertebrate eye, for which we have, according to (3)

$$w = 2{\cdot}1\ \lambda/d = 3\ \lambda/R \qquad \ldots.(8)$$

assuming $d = 0{\cdot}7R$ which corresponds to a "stop" of 3 (see above). It is possible that the vertebrate eye never attains this theoretical limit with this wide pupil—at least not in the case of the human eye, where w is about twice the value given by (8). Comparing (7) and (8) we find

$$(w_{\text{insect}}/w_{\text{vertebr.}})^2 = q^2 = 0{\cdot}08\ R/\lambda \qquad \ldots.(9)$$

For a large eye, with R say 1 cm, the construction of the insect eye is obviously far inferior to the vertebrate eye, q being 40 for $\lambda = 0{\cdot}5\mu$. For smaller eyes, with a radius of say 0·2 mm q is still 8, or 4 if these small vertebrate eyes also do not reach the limit given by (8). Since the (angular) errors due to spherical aberration, etc., do not increase with decreasing dimensions of the eye, they become relatively less important than diffraction errors and it is probable that (8) is correct for the small vertebrate eyes. We can therefore conclude that the insect eye is inferior to the vertebrate eye—unless its better "temporal" resolution has an appreciable survival value. It is evident that, once it had been "decided" that the eye of insects should consist of a group of partial eyes looking in different directions, nature has made the best of it. And we should not forget that an eye is not of any value without nerve-circuits. With the same radius R the vertebrate eye contains more sense cells. Consequently also a more bulky nerve system is required, unless the nerve system is really adequate to give the maximum performance only for part of the retina (in the human eye the fovea).

§ 4. *The sense cells as waveguides*

The rods and cones in the vertebrate retina have a higher refractive index than the surrounding media, consequently light will be totally internally reflected at the boundaries of the rods and cones (see Fig. 22)

if it does not diverge too much from their axes. As a result of these reflections radiation is captured by the cell, even though the cell is long relative to its diameter. BRÜCKE (cited by ref. [52]) suggested this mechanism more than a century ago. In the sense cells of the insect eye we have a similar effect. The cell is embedded in black pigment and consists of a "thread" of high refractive index (about 1·5) (the rhabdomer), surrounded by the fluid of the cell which has a lower refractive index (about 1·33). Consequently light will be captured by the rhabdomer. It is probable that the photochemical substance is *also* concentrated in the rhabdomer. In the eye of hymenoptera (e.g. the

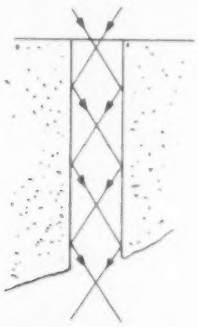


Fig. 22. Collimation of a beam of light total reflection in a "tube" with higher refractive index.

bee) the cells of one partial eye form a cylinder (surrounded by pigment) and the corresponding rhabdomers together form one large central rhabdom.

Depending on the refractive indices, the internal reflection decreases with increasing angle between the incident beam and the axis of the cell. Consequently, the sensitivity of the cell will decrease with increasing divergence. This directional sensitivity has been found for the human retina; it is the so-called STILES-CRAWFORD effect (ref. [51]). These authors found that the threshold of vision increased if the (narrow) incident beam was displaced from the centre of the pupil towards its edge. In the latter case the beam will fall obliquely on the retina.

The explanation of the STILES-CRAWFORD effect was found some years after its discovery and instead of the elementary explanation by geometrical optics a more rigorous explanation is obtained by applying the theory of waveguides and dielectric antennae to these small structures. We refer to BRIAN O'BRIEN[52] who made a large model of the sense cell where short (radar) waves were used instead of visible light and to a paper by TORALDO DI FRANCIA.[53]

§ 5. *The detection of polarized light*

In recent years it has been found that various insects (more generally arthropods) can detect the plane of polarization of polarized light. It was first shown by VON FRISCH (ref. [54]) for the bee and subsequently for other insects (e.g. ants). This ability is of great practical use to these animals, since it allows them to determine the position of the sun even if it is hidden behind clouds. The state of polarization of the light from any piece of the blue sky left gives information about the position of the sun. The position of the sun, is the point of orientation

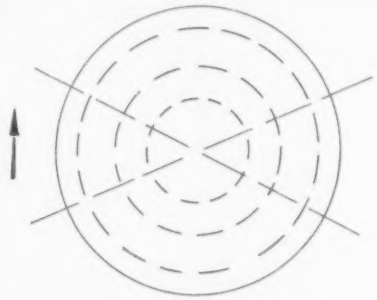


Fig. 23. Arrangement of absorbing molecules with circular symmetry (see text).

during long flights (or walks). What interests us here, is mainly the way in which this polarization is detected. For all vertebrates studied, no such ability to determine the plane of polarization of light has been found—that is if the entoptic phenomenon in the human eye, known as the "Haidinger" brushes, is disregarded. HAIDINGER discovered in 1844 that if one looks through a nicol prism (or polaroid) at a white field one sees a yellow figure. The figure can be described as a figure 8 perpendicular to the plane containing the electric vector (direction of polarization). The explanation of the phenomenon was given by HELMHOLTZ, who assumed that the molecules in the yellow macular pigment were orientated with radial symmetry, i.e. perpendicular to the fibres which radiate outward from the fovea (see Fig. 23). If the incident light is polarized in the direction of the arrow, the pigment in the segment enclosed by the dotted lines will absorb more of the light than do the other parts. Consequently one will see a dark brush if the incident light is blue, and a yellow figure if it is white. A few years ago[49], [55] we confirmed this explanation by quantitative measurements of the effect. It was found that it had the same spectral distribution as the absorption curve of the pigment. Moreover it was found that only about 5 per cent of the molecules were orientated—a reasonable value since nature has not orientated the molecules "on purpose."

Recently NAYLOR and STANWORTH[56] have carefully explored the magnitude of the effect in smaller test fields in the fovea.

It is obvious that the detection of polarization of light by insects is a much more fundamental phenomenon than is the Haidinger effect for human vision. It was soon found that this ability to detect polarization was not due to the presence of "polaroid" filters either in the lenses or elsewhere in the optical path. (See AUTRUM and STUMPF.[57]) If the material of the insect eye is observed through a rotating nicol prism, no obvious changes in the intensity of the transmitted light are found—and if *we* cannot perceive any difference of intensity it is improbable that there is any dichroic effect present which is of importance for the insect.

A very ingenious theory was suggested by MENZER and STOCKHAMMER, which was based on their detection of double refraction in the sense cells. The refractive index for a beam of light that traverses a sense cell along its axis depends on the plane of polarization. The "optical axes" in a section perpendicular to the length of the cell are different, in a typical way, for the 7 cells of one partial eye. (Various insects were studied, among others calliphora.) The difference between the two refractive indices was found to be about 0·001 at the most. We have mentioned earlier that the light received by the eye is mainly concentrated in the rhabdomer. MENZER and STOCKHAMMER assume that the rhabdomer consists of two parts, a doubly refracting central column surrounded by a cylinder with a refractive index intermediate between the two refractive indices of the central cylinder. In this system the extraordinary component would be concentrated in the central cylinder, whereas the ordinary component could escape into the outer ring. If only one of these regions contains the photochemical substance, the eye would be sensitive to one component of the incident light only. We might, however, object that because of the small difference in refractive index total reflection will only occur if the beam does not diverge more than 2° from the axis. (See also ref.[52], [53].) The actual aperture of the cone of light that illuminates the cell is however, much larger than 2°. We have repeated the measurements of MENZER and STOCKHAMMER, and we could not find any double refraction. During a visit to their laboratory, it was discovered that the double refraction was found in eyes fixed by alcohol. They agreed that fresh eyes did not show double refraction*. At any rate the double refraction is weaker than we assumed above, so that the permissible divergence of the beam from the axis is still smaller than 2° and consequently the mechanism could not be very effective.

* According to a private communication (in 1951), STOCKHAMMER has finally found a weak double refraction in fresh eyes.

For our experiments with the insect eye we had another working hypothesis. It is well known (see SCHMIDT[59]) that in the vertebrate eye the molecules of the photochemical substance are arranged in such a way that their axes of maximum absorption are perpendicular to the long axis of the cell. This is obviously the optimal arrangement since the incident light only has a component of the electrical vector perpendicular to the axis, i.e. parallel to the molecules. There is no special order *in* planes perpendicular to the length of the cell; on the contrary, we obtained experimental evidence that in these planes the molecules can move freely.[49], [60] Considering now the insect eye it is obvious that an ability to detect the plane of polarization would result if all the molecules of the photochemical substance were oriented the same direction. The cell would only respond to the component of the light vibrating in the same direction as the molecules.

We have tried to check this hypothesis* by investigations of the eyes of some dipters, mainly calliphora. We have already mentioned that we did not find any double refraction of the cells. Our search for a selective absorption of any one direction of polarization also produced a negative result, even though a difference of less than one per cent could have been detected. Since the work has not yet been published, we give some experimental data.

The anatomical technique was the same as that used by MENZER and STOCKHAMMER (l.c.). From intact eyes we cut, with a freezing microtome, slices that consisted of intact lenses, etc., and nearly the whole length of the sense cells. These slices were mounted on the stage of the microscope so that the light entered and passed through the eye in the normal way until it reached the plane of cutting; the microscope was focused on this plane. We started by taking microphotographs for different directions of polarization of the incident light, and we found that no variations in the relative brightness of the cells of one ommatidium could be detected. This procedure was not very accurate, however. In further experiments the microscope was used as a "projection microscope." Through a hole in the projection screen the light from only one receptor was allowed to fall on a photomultiplier. We never found any effect produced by turning a polaroid in the incident beam, even though a large number of eyes, generally dark adapted, were used in the work. The use of a projection microscope suggests that the eye was illuminated with an intense light; the intensity used was only just sufficient, however, to see the projected image with a dark adapted eye in Maxwellian view. The photometric arrangement used was the optimal. For some of the measurements chopped light was used, so

* This work on insect eyes was done during the summer of 1951 by the biophysical group in our laboratory and I might especially mention the work of J. W. KUIPER, (biol. drs).

that variations in the dark-current of the multiplier did not interfere and the accuracy was further improved by rotating the polaroid 50 times per second. Any dichroism of the eye would have produced a response of the multiplier with a frequency of 100 cycles/sec—but again no effect was obtained. Generally blue light was used.

This result does not disprove our theory it only means that the absorption by the photochemical substance is less than one per cent. This may perhaps seem to be a very low efficiency for the receptor, but it should be borne in mind that although the efficiency of the human dark adapted rods is about 15 per cent (see RUSHTON[61]), no absorption by photopigments in the cones has yet been demonstrated. Finally, we emphasize that the double refraction found by MENZER and STOCKHAMMER (l.c.) even though it was found in fixed material, proves that an anisotropic arrangement of molecules is present. This obviously agrees with our hypothesis; the only missing link is the proof that the molecules of the photochemical substance are themselves oriented, in this anisotropic "matrix."

Using the experimental arrangement of the insect eye, described above, we could also study the directional sensitivity of individual cells. The microscope was used without a condenser; the insect eye could now "look" at a small light source which could be moved about in its visual field. Looking through the microscope we could see whether light really came through a given cell. It was found that the point source had to be displaced through an angle (seen from the eye) of 4° from the moment that the cell became illuminated until it became nearly dark again. This means that the visual field of one cell is 4°. This agrees very well (we observed with our own eye) with the theoretical prediction in (1), inserting for λ $0{\cdot}5^5\,\mu$ and for the diameter of one lens $30\,\mu$, and taking into account the diameter D of the rhabdomer. If the light source was displaced further, diffracted light from the first fringe of the Airy-pattern became visible through the cell. The intensity of this fringe is much less than for the central maximum. It is most important that adjacent cells in one ommatidium do not shine up at the same time, if the light source is displaced. Obviously their visual fields are different. They overlap in such a way that the centre of one field is just on the edge of its neighbour. This important result was anticipated already on page 249. The field of one ommatidium, disregarding the weaker diffraction fringes around the central maximum, has consequently a diameter of about 8°.

So far we have confined ourselves to the eyes of dipters which have seven separated cells. We have already mentioned, however, that the rhabdomers in the eye of hymenoptera (e.g. the bee) grow together to form one central rhabdom. If absorption of light in one segment of

this rhabdom is still the stimulus for the cell to which this segment originally belonged as the rhabdomer, then this arrangement, at least if our explanation of the effect is right, does not interfere with the detection of the plane of polarization. Another question, however, is whether the seven cells still have different visual fields. If there is no "screen" between the original segments, i.e. if the rhabdom is optically a unit, each ommatidium will only detect one "point" in the visual field. A layer with a refractive index very different from the refractive index of the rhabdomer itself, could, by the total reflection it produces, act as a screen and its thickness need only be a fraction of a wavelength (the actual thickness depending on the difference in refractive indices). The only way to answer this question is to perform the same experiment as described above for the dipterian eye. Our first experiments were not conclusive since the eye is very difficult to cut, the lenses and the crystal cone being very hard. For another type of observation, one made by ANTHONIE LEEUWENHOEK in the seventeenth century, this rigidity is very convenient since it makes it possible to remove from the eye all soft material, such as sense cells and pigment, so that only the dioptric apparatus is left. In the focal plane of each ommatidium one then sees the image of suitable objects such as windows, or if the light source is small the diffraction patterns discussed on page 246. We have checked that the diameters of the fringes observed agree with the theory, which is not remarkable.

VOL
6
195

§ 6. *Lightquanta and vision*

This aspect of vision is at present perhaps the most interesting. The fundamental developments date from the period 1942–45, and have come independently from the Netherlands (DE VRIES[62], VAN DE VELDEN[63]) and from the U.S.A. (HECHT *et al.*[64]). The main line of thought has been discussed by PIRENNE[45] in a previous issue of *Progress in Biophysics*. We shall confine ourselves to some related problems, and even then we cannot include all the recent developments not discussed by PIRENNE. For the work of BOUMAN and other recent developments, we refer to [66].

One interesting problem is the question whether the rods, which respond to a single quantum of light, do this in a way similar to Geiger counters. In our paper[62] we gave evidence that the reaction of the rods is more complex. A counter gives one standard reaction, irrespective of the number of quanta absorbed in the reaction-time. According to HARTLINES' work[67] on Limulus, the nerve fibres at least give a more complex response to short flashes which consist of more than one discharge. Perhaps the cells themselves do not give a discrete response at all, but an effect similar to the microphonic effect of the labyrinth (here the electroretinogram).

More important is the problem of *adaptation*. It is well known that the sensitivity of the eye decreases if it is exposed to bright light. This phenomenon is called adaptation since it results in a better performance at the higher intensity, as is also the case for adaptation to low intensities. Adaptation to bright light is accompanied by a bleaching of the retina, i.e. a decrease of the amount of the photochemical pigment (rhodopsin) in the *rods*. Nothing is known about the corresponding process which occurs in the cones. Before the war, HECHT had developed his well known photochemical theory of vision, claiming a close relation between the concentration of rhodopsin, or its products of photodecomposition, and the visual process. Since it has been proved that one rod can be triggered by one quantum of light, this macro-chemical theory has lost its basis. The mass-laws of chemical reactions cannot be given much credit if only one molecule is present in the cell—and even at the brightness of a just sufficiently illuminated room the number of quanta absorbed per second and per rod are barely a few hundred (DE VRIES[62]).

There is another more interesting difficulty. The photochemical theory relates the sensitivity of the eye to the concentration of the rhodopsin. It is generally accepted at present, however, that the increase of the threshold is already appreciable for intensities and brightnesses of the adapting light which hardly produce any bleaching. The increase of threshold is measured experimentally and the bleaching is calculated (for a first discussion of the problem and the calculation see ref. [62] page 555 and ref. [67]). To calculate the rate of bleaching of the rhodopsin the effective cross section of the molecule is required; this cross section is known from *in vitro* experiment on the bleaching of rhodopsin solutions. One could of course object that the rate of bleaching in the retina might be higher because of secondary processes, but some observations on a fresh frog retina produced no evidence of such effects. More recently RUSHTON[61] has given the whole discussion an exact basis by measuring the amount of rhodopsin in the intact human eye during adaptation. He found that the rate of bleaching was the same as in the *in vitro* experiments.

ROSE[65] also came to the conclusion that adaptation is not due primarily to the bleaching of rhodopsin by the following reasoning. The quantum character of light involves a fundamental limitation in the minimum perceptible difference ΔI in the brightness I of two adjacent test fields. If the retina receives from one test field on the average, N quanta which are effectively absorbed, the fluctuations are of the order of $\sqrt{N}$. If the other field has about the same brightness, then the fluctuations in the difference ΔI are of the order of $\sqrt{2N}$, consequently it will be impossible to detect a difference between the two fields that does not exceed these fluctuations by a certain factor k.

It is possible to compare this theoretical limit with the actual performance for a given intensity I by first calculating N from I. This calculation is not very exact since the fraction f of the quanta which is effectively absorbed by the retina is not known exactly. Within the limits of error, however, $\Delta N/N$ is of the same order of magnitude as $\Delta I/I$ determined by experiment. This means that the eye practically reaches the theoretical limit. This is not only true for one intensity, but holds for a wide range of intensities, since $\Delta I/I$ decreases with increasing brightness in proportion to $\sqrt{I}$. Only when the brightness is so high that the difference ΔI is about 1 per cent of I, does $\Delta I/I$ not decrease further (obviously by physiological limitations). So far the results obtained by DE VRIES (l.c.) and ROSE (l.c.) are the same. ROSE, however, points out that the fact that the change of $I/\Delta I$ is proportional to $\sqrt{I}$ implies that the efficiency of the retina is constant. Consequently there cannot be any appreciable bleaching, nor can any other processes occur that affect the efficiency of the retina. The constant efficiency of the eye has recently been proved by BOUMAN[73], by similar reasoning, but with special adequate experiment.

WALD[68] has suggested the following solution for the discrepancy between the rise of the threshold and the correspondingly very small bleaching of the rhodopsin (at least at lower intensities; on a bright summer day, of course, the bleaching is very important). It should first be mentioned that according to electron microscope studies the rod is subdivided into thin discs. The fundamental assumption is that a "hit" on one molecule of rhodopsin inactivates the whole "disc" of which it is a part. As a consequence of this the decrease of sensitivity would be larger than the degree of bleaching. According to the preceding paragraph, however, this hypothesis would imply an intensity discrimination much worse than it actually is, since it postulates a decrease of the efficiency with increasing intensity.

Obviously adaptation does take place on another level of the visual process, e.g. in the receptor itself, where the effect of one quantum of light is amplified to a macro-reaction of the cell, or in the layer of nerve cells in the retina. Compared with a radio receiver (ROSE[65]), *these* adapting mechanisms are equivalent to a control of the gain of the amplifier in one of its last power stages, whereas bleaching can be compared to a shortening of the antenna. It is obvious that the first method is by far the best because it also reduces the noise.

§ 7. *Effects of thermal agitation*

The first effect of thermal agitation might be *dissociation* of the photochemical substances, or whatever is the effect produced by the light quantum. The probability that this dissociation occurs, is according to section IV, 4, $f \exp(-E/kT)$. The mean life time of a certain state

in a fluid or solid is about 10^{-13} sec.; inserting for f the factor 10^{13} we obtain the well-known formula for the rate of a monomolecular reaction. Since one rod contains about 10^9 colour groups of rhodopsin (ref. [43]), the probability that a rod is excited is $10^{22} \exp(-E/kT)$ per second. The permissible number of spurious excitations is somewhat arbitrary. One must consider the fact that a sensation of light occurs, only if two

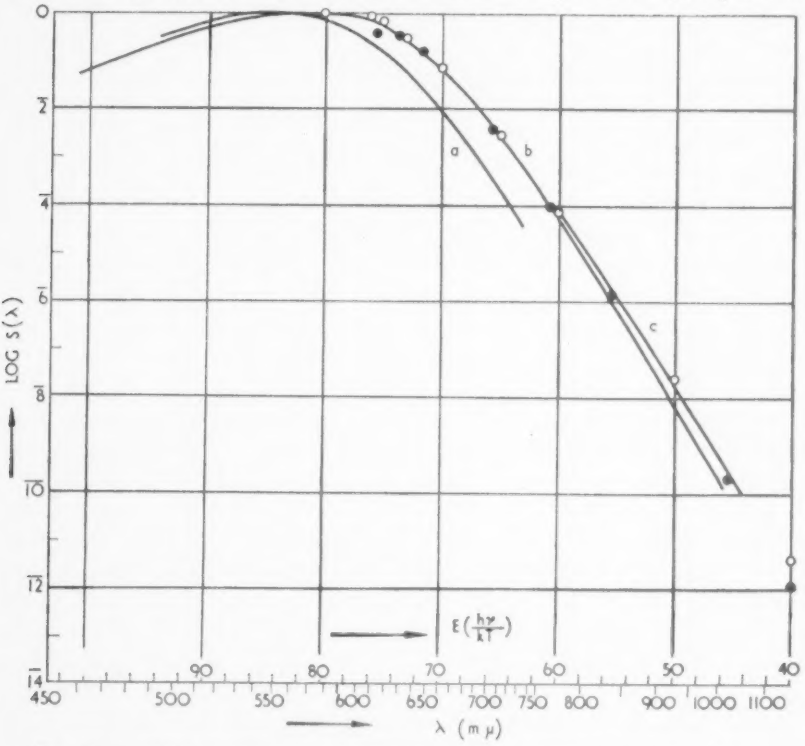


Fig. 24. Sensitivity curves of red receptors (b) and green receptors (a) of the human eye (from ref. 69, 71). The horizontal axis gives the energy of the quanta, expressed in kT. ○ and ● calculated from (10) for $s = 6$ and $s = 4$ respectively. Curve c: part of the absorption curve of the rods. The experimental data between 75° and 1050 mμ from 74.

rods, in a group of about 100 rods, are excited within 0·02 sec. (VAN DE VELDEN[63] and PIRENNE[45]). This coincidence-mechanism suppresses some of the spurious effects. There are about 10^6 of these receptor groups in the retina. If we assume that ten of these spurious elementary light sensations can be allowed per second one easily calculates that the probability of a spurious effect per rod should be less than 0·0003, and that the energy E required for decomposition of the rhodopsin should be at least $58\,kT$. This energy corresponds to a quantum of light with a wavelength of 800 mμ. The maximum of the absorption curve of

rhodopsin is at an even shorter wavelength (505 mμ), so that the spurious sensations produced by thermal agitation will not be disturbing. One can conclude that a photochemical substance that can be dissociated by infra-red light will have a dissociation energy so low that spurious sensations could only be avoided if its concentration were very low, and this implies a very low sensitivity. (See also DE VRIES.[31], [43], [69])

Thermal agitation can also play a more positive role when the light that falls on the retina has not enough energy to decompose the photochemical substance, i.e. $h\nu < E$. In this case the thermal agitation can provide the missing amount of energy W. The probability that the molecule (or the colour group) has this energy if only one degree of freedom plays a part is $\exp(-W/kT)$. If more degrees of freedom have to be considered, the expression is somewhat more complicated, namely:

$$w = \sum_{r=0}^{s-1} \frac{1}{r!} (W/kT)^r e^{-W/kT} \quad \ldots\ldots(10)$$

The exponential term is still dominant, especially if W/skT is larger than 5 or 10. In this region the probability of absorption, or decomposition, decreases by a factor of e if $h\nu$ decreases by kT. Indeed the sensitivity curves of Fig. 24 demonstrate this slope, disregarding some small deviations which have to be ascribed to individual properties of the molecules. That these will play a role becomes clear if we consider the absorption in more detail.

Fig. 25 shows the well-known Franck-Condon diagram. The horizontal axis gives the value of a parameter which in case of a diatomic molecule is the distance apart of the atoms. For a polyatomic molecule the parameter is one of the more generalized coordinates which describe the configuration of the atoms. The ordinate axis gives the potential energy. Curve I shows the potential energy for the ground state of the molecule, curve II for an excited state with a different electronic configuration (schematical).

If the molecule is in a lower state, it vibrates with a small amplitude from A to A''. If it has an extra vibrational energy W it vibrates from B to B''. The FRANCK-CONDON principle says that electronic transitions generally start from the extreme positions A or A'', etc., since according to the classical interpretation the molecules stay a longer time at these points. The transition A to A' requires more energy than does that from B to B', because the molecule at B has already the thermal energy W. In case a, however, the difference is much larger than W, because of the vertical distance W' between A' and B'. In the elementary discussion of the previous paragraph we have tacitly

assumed that situation *b* was correct; unfortunately a horizontal curve II can only be an exception. (See also below, small print.) If the molecule has an energy W, the quantum can be W smaller and not more. This implies:

(*a*) The slope of the sensitivity curve (Fig. 24) is a factor e for a variation of the quantum energy by kT;

(*b*) If T varies, the slope of the curve changes proportionally.

In situation *a* the slope will be much smaller, since the quantum energy can decrease more than the energy that is provided by thermal

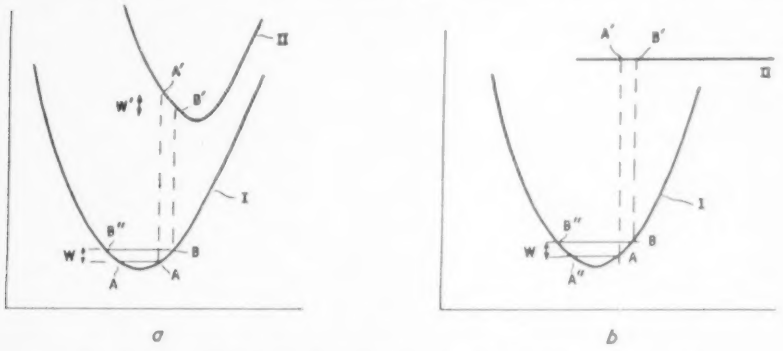


Fig. 25. Franck-Condon diagrams (*a*) and (*b*) for different potential curves of the second energy level II. Schematical.

agitation. Furthermore the broadening of the curve is no longer proportional to T but to $\sqrt{T}$. The energy W increases in proportion to T; the amplitude of the vibration AA' is proportional to $\sqrt{W}$ and it is this horizontal displacement, multiplied by the slope of curve II that governs the decrease of the distance AA'. This proportionality between the width of the absorption curve and $\sqrt{T}$ is well known for the F-centres in alkali-halogenide crystals (see MOTT[70]).

In order to find an argument why case *b* holds for the photochemical substances, we have studied some dyes in solution at different temperatures*. The substances used were potassium permanganate, potassium chromate and bichromate; diacetyl (dissolved in Xylol), chrysasin (dissolved in tetrahydrofurfuryl alcohol) and β-carotine (in Xylol). The first group of three substances showed the properties correlated with case *a*, the second group belonged obviously to case *b*. It is not satisfactory that this result could not be predicted. β-carotine had a slope of e per kT only at about 540 mμ. At longer wavelengths the curve had again a hump; here no clear effect of temperature was found.

* Together with J. DE BOER (unpublished).

Perhaps the background of the behaviour is that in the first group only a small number of atoms play a part in the absorption of light, whereas many atoms are involved in the second group. So we understand first that the equilibrium position in the excited state will be displaced more in the first group than in the second group. Then the slope of II above A and B will be much less than in Fig. a because of the smaller horizontal shift. Even if II is displaced, relative to I, we have to consider more coordinates in a more complicated molecule, each with a diagram as in case a. Up to now we could say that one-half of the molecules are in the favourable position A or B (the other half being in A'' or B''). If there are a large number of coordinates, each having an energy W ($\approx kT$), they will rarely be all at the same time in their favourable position, so that the extra gain W' becomes less important than the sum of all thermal energies.

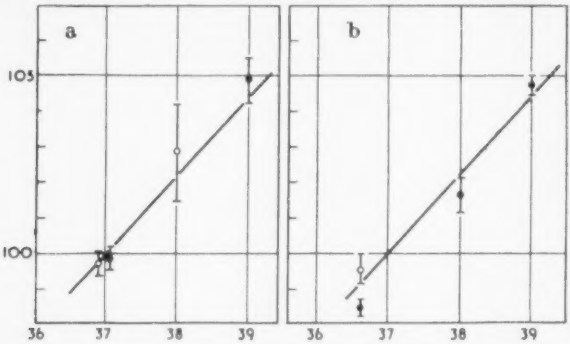


Fig. 26. Effect of the temperature of the observer on a brightness match between two wavelengths (660 and 730 mμ) for two different observers. (a) Deuteranomalous observer; (b) deuteranope. Lines calculated from the theory. ○ rising temperature. ● decreasing temperature (from ref. 72).

The shape of the red tail of the sensitivity curve of the eye became of interest during the course of an investigation of colour division (DE VRIES[71]). The basic fact was that normal as well as abnormal and even colourblind observers accepted the same brightness matches in the red region of the spectrum even though the absolute sensitivities of these observers were very different. This finding means, that on a logarithmic scale, the sensitivity curves are parallel. This could be explained by the "physiological" assumption that all eyes contained the same receptor for red, but in different amounts. The "physical" explanation given above, however, was more attractive, even from the point of view of colourvision.

Another check on this theory was the demonstration of the influence of temperature on the sensitivity curve in the red region (DE VRIES[72]). The temperature of the eye was changed by putting the observers in a hot bath. Fig. 26 gives the influence of temperature on a brightness match between two different red test fields; one being illuminated by 660 mμ, the other by 730 mμ. At higher temperatures of the eye less of the longer wavelength was required in order to make the brightness match with that of the shorter wavelength. The lines in Fig. 26 were calculated on the assumption that the slope of the curve on a logarithmic scale (Fig. 24) was proportional to $1/T$. The agreement is excellent.

References

(1) Proetz, A. W.; *Applied Physiol. of the Nose.* Annals Publishing Company, St. Louis, 1953

(2) Zwaardemaker, H.; *Hdb. Hals, Nasen usw. Heilkunde*, v. Denker u. Kahler (Bd. I)

(3) Davies, J. T. and Taylor, F. H.; *Nature, Lond.* 74 (1954) 693

(4) Neuhaus, W.; *Z. Vergl. Physiol.* 35 (1953) 527

(5) Dijkgraaf, S.; *Experienta* 8 (1952) 205

(6) de Vries, Hl.; *Acta Oto-Laryng.* 38 (1950) 262

(7) Pumphrey, R. J.; *Symp. Soc. Exp. Biol.* IV, 1950

(8) Frisch, K. von; *Nature, Lond.* 41 (1938) 8

(9) Schneider, H.; *Z. Vergl. Physiol.* 29 (1941) 172

(10) Dijkgraaf, S.; *Z. Vergl. Physiol.* 28 (1941) 389

(11) Löwenstein, O. and Roberts, T. D. M.; *J. Physiol.* 114 (1951) 471

(12) Autrum, H.; *Naturwiss.* 38 (1951) 434

(13) Oettinger, H.; *Zool. Jahrb. Abt. Alg. Zool. Physiol.* 58 (1937–38) 181

(14) Wulf, B.; *Arch. Anat. u. Entw. gesch. Anat. Abt.* 57 (1901)

(15) Doesschate, G. ten; *Onderz. Physiol. Lab. Utrecht*, 5° reeks, 14 (1914) 1–75

(16) Steinhausen, W.; *Pflüg. Arch.* 228 (1931) 332

(17) Graybiel, A.; *Amer. J. Psych.* 61 (1948) 21

(18) Groen, J. J.; *Practica Oto-Rhino-Laryng.*, Vol. 14, Suppl. 2, page 57

(19) Hallpike, C. S. and Hood, J. D.; *Proc. Roy. Soc.* B 141 (1953) 216

(20) Békésy, G. von and Rosenblith, W. A.; *Handbook of Experimental Psychology*, John Wiley & Sons, 1951

(21) Békésy, G. von; *J. Acoust. Soc. Amer.* 19 (1947) 452

(22) *Ibid.* 21 (1949) 233

(23) *Ibid.* 21 (1949) 245

(24) Zwislocky, J.; *Acta Oto-Laryng.* Suppl. LXXII (1948) 53

(25) Huggins, W. H.; Thesis, Cambridge (Mass.) 1953

(26) Tröger, J.; *Physik. Z.* 31 (1930) 26

(27) Békésy, G. von; *J. Acoust. Soc. Amer.* 21 (1949) 217

(28) Wever, E. G. and Bray, C. W.; *J. Exp. Psych.* 13 (1930) 373

(29) Humphrey, R. J.; *Nature, Lond.* 143 (1939) 1

(30) Vries, Hl. de; *Physica* 14 (1948) 48

(31) Vries, Hl. de; *Experientia* 4 (1948) 205

(32) Vries, Hl. de, Jielof, R. and Spoor, A.; *J. Physiol.* 116 (1952) 137

(33) Vries, Hl. de and Bleeker, J. D. J. W.; *Acta Oto-Laryng.* 37 (1949) 290

(34) Vries, Hl. de, Kuiper, J. W. and Ubbens, D. R.; *Nature, Lond.* 175 (1955) 303

(35) Békésy, G. von; *J. Acoust. Soc. Amer.* 23 (1951) 576

(36) Holst, E. von; *Z. Vergl. Physiol.* 32 (1950) 60

(37) Vilstrup, T.; *Structure and Function of the Semicircular Canals*, Ejnar Munksgaard, Copenhagen, 1950, page 39

(38) Vries, Hl. de and Vrolijk, J.; *Acta Oto-Laryng.* 43 (1953) 80

(39) Schmaltz, G.; *Proc. Roy. Soc. Med.* 25 (1931) 360

(40) Vries, Hl. de; *J. Acoust. Soc. Amer.* 24 (1952) 527

(41) Sivian, L. J. and White, S. D.; *J. Acoust. Soc. Amer.* 4 (1933) 288

(42) Vries, Hl. de; *Acta Otol-Laryng.* 36 (1948) 230

(43) Vries, Hl. de; *Nature, Lond.* 161 (1948) 63

(44) Newman, E. B.; *Ann. Rev. Psych.* 1 (1950) 49

(45) Pirenne, M. H.; *Progress in Biophysics and Biophysical Chemistry* 2 (1951) 193

(46) MALLOCK, A.; *Nature, Lond.* 110 (1922) 771

(47) VRIES, HL. DE; *Documenta Ophtalmologica* III, (1949), 92

(48) EXNER, S.; *Physiol. der fazettierten Augen v. Krebson u. Insekten*, Leipzig-Wien, 1891

(49) VRIES, HL. DE, JIELOF, R. and SPOOR, A.; *Physica.* 19 (1953) 419

(50) AUTRUM, H.; *Z. Natur. forsch.* 5*b* (1950) 38

(51) STILES, W. S. and CRAWFORD, B. H.; *Proc. Roy. Soc.* (London) B 12 (1933) 428 and B 116 (1934) 55

(52) O'BRIEN, BRIAN; *J. Opt. Soc. Amer.* 39 (1949) 12, and 41 (1951) 888

(53) TORALDO DI FRANCIA, G.; *J. Opt. Soc. Amer.* 39 (1949) 324

(54) FRISCH, K. VON; *Naturwiss.* 35 (1948) 12

(55) VRIES, HL. DE, JIELOF, R. and SPOOR, A.; *Nature, Lond.* 166 (1950) 958

(56) NAYLOR, E. J. and STANWORTH, A.; *J. Physiol.* 124 (1954) 543

(57) AUTRUM, H. and STUMPF, H.; *Z. Naturforsch.* 5*b* (1950) Hft. 2

(58) MENZER, G. and STOCKHAMMER, K.; *Naturwiss.* 38 (1951) 190

(59) SCHMIDT, W. J.; *Z. Zellforsch.* 22 (1935) 485

(60) VRIES, HL. DE, JIELOF, R. and SPOOR, A.; *Physica* 19 (1953) 419

(61) RUSHTON, W. A. H. and HAGINS, W. A.; *J. Physiol.* 120 (1953), 61P and *Nature, Lond.* 174 (1954) 1096

(62) VRIES, HL. DE; *Physica* 10 (1943) 553

(63) VELDEN, H. A. v. D.; *Physica* 11 (1944) 179

(64) HECHT, S., SHLAER, S. and PIRENNE, M. H.; *J. Gen. Physiol.* 25 (1942) 819

(65) ROSE, A.; *J. Opt. Soc. Amer.* 38 (1948) 196

(66) BOUMAN, M. A.; *J. Opt. Soc. Amer.* 45 (1955) 36

(67) VRIES, HL. DE; *Physica* 14 (1948) 319 (see page 345–348)

(68) WALD, G.; *Science* 119 (1954) 887

(69) VRIES, HL. DE; *Ned. T. Nat.* 13 (1947) 191

(70) MOTT, N. F. and GURNEY, R. W.; *Electronic Process in Ionic Crystals*, Clarendon Press, 1940

(71) VRIES, HL. DE; *Physica* 14 (1948) 367

(72) VRIES, HL. DE; *Experientia* 4 (1948) 357

(73) BOUMAN, M. A.; *J. Opt. Soc. Amer.* 42 (1952) 820

(74) GRIFFIN, D. R., HUBBARD, R. and WALD, G.; *J. Opt. Soc. Amer.* 37 (1947) 546; see also GOODEVE, C. F.; *Proc. Roy. Soc.* A 155 (1936) 664

(75) BARLOW, H. B.; *J, Exp. Biol.* 29 (1952) 667–674

VOL.
6
956